2015

The Social Determinants of Success: The Effect of Status, Framing, and Group Membership on Test Scores

Jennifer L. McLeer
University of South Carolina - Columbia

THE SOCIAL DETERMINANTS OF SUCCESS: THE EFFECT OF STATUS, FRAMING, AND GROUP MEMBERSHIP ON TEST SCORES

by

Jennifer Lynn McLeer

Bachelor of Arts
Manhattan College, 2007

Master of Arts
University of South Carolina, 2009

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Sociology

College of Arts and Sciences

University of South Carolina

2015

Accepted by:

Barry Markovsky, Major Professor

David Melamed, Committee Member

Shane Thye, Committee Member

Michael Lovaglia, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

ACKNOWLEDGEMENTS

This research was supported by a National Science Foundation Dissertation Improvement Grant (SES-1029741) and the University of South Carolina's Dean's Dissertation Fellowship. I also received many helpful comments over the course of this project from Barry Markovsky, David Melamed, Brent Simpson, Shane Thye, Michael Lovaglia, and many of my graduate student colleagues.

ABSTRACT

Previous findings have shown that intelligence test scores of members of disadvantaged groups are negatively affected by their relative status in test situations. We explain this by assuming that lower-status actors forego benefits that normally follow from successful performance of a task if they anticipate that there are also costs associated with success. We use an integration of Status Characteristics Theory, Rational Choice Theory, Prospect Theory, and Self-Categorization Theory to argue that compared to high status individuals, those of lower status should be particularly affected by costs associated with high performance. We ran an experiment that manipulated factors such as perceived social status and rewards/costs associated with intelligence test performance. We found that low status individuals do, in fact, score lower than high status individuals when they anticipate costs for high test performance. We further identified that the underperformance was due to a conscious attempt to put forth less effort on the test. A follow-up study assessed the extent to which the social status of the person giving the sanctions contributes to this relationship. We found that low status individuals are particularly affected by sanctions that their in-group give for high performances. We discuss the implications of these findings in the context of educational inequality.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Intelligence test scores set off a domino effect of life outcomes for the test-taker. Those with higher intelligence test scores tend to become more educated, get higher paying jobs, and have more prestigious occupations (Collins 1979; Jencks et al.1972; Jencks et al.1979; Hauser et al. 2000; Grodsky and Pager 2001; Hall and Farkas 2011). High scorers are also more likely to contribute to society via productive economic behavior, civic engagement, and a decreased rate of criminal activity (Nisbett 1998; Putnam 1995). On the whole, higher intelligence test scores give individuals many of the pieces necessary to have successful life outcomes.

And yet, the scores of entire categories of people are artificially decremented by social processes. Those from a high socioeconomic background consistently outperform those from more disadvantaged backgrounds on a variety of educational indictors (Blau and Duncan 1967; Bordieu 1977; Hauser et al. 1983; Sewell et al. 1969). Racial differences in test scores are also widely prevalent (Nisbett 1998). Given that intelligence tests are highly consequential and that there are categorical differences in scores, we must be certain that scores reflect a test-takers actual intellectual abilities and are not capturing the remnants of some other process. However, differences in test scores are not solely due to differences in academic abilities. Individuals are susceptible to the effects of social processes, such as stereotype threat, socially motivated underachievement, and status-related bias (Steele and Aronson 1995; White et al. 2002; Lovaglia et al. 1998). As a

result of these social impediments, many individuals score lower on intelligence tests than their abilities warrant, and thus position themselves on a track to gain less individually and contribute less collectively.

The fact that these social processes are at play, in general, is problematic. However, the true danger lies in the fact that the same categories of people that tend to score lower on intelligence tests are the ones that are differentially affected by the social pressures to underachieve. On the whole, racial minorities and low-income students tend to have more experiences that encourage them to underinvest in education, such as norms of disengagement with school, peer-to-peer sanctioning for high achievement, and familial dissuasion from long term educational goals (Fordham and Ogbu 1986; Fryer and Torelli 2010; Austen-Smith and Fryer 2005). These experiences are unparalleled in racial majority and high-income environments.

Taken together, these propositions not only showcase how easy it is for a disadvantaged individual to remain fixed within a lower social class than their intelligence warrants, but they also highlight a real societal level problem. As it stands, we are doing little to address a major driving force behind social inequality that could significantly improve mobility between classes. Taking measures to identify and understand the negative consequences of the social processes behind underachievement is the first step in rectifying this situation.

To help understand how social factors affect performance on mental ability tests, this study draws on four literatures connected to the study of group processes: Status Characteristics Theory, Rational Choice Theory, Prospect Theory, and Self-Categorization Theory. We use these theories to support an argument concerning the

rewards/costs associated with performance on academic tests. We are interested in why members of disadvantaged groups such as racial minorities and low-income students tend to under-perform on such tests, even in situations where their actual abilities exceed those of the majority group members.

We offer four main contributions here. First, we test the internal validity of the argument that those of disadvantaged status score lower on intelligence tests than their advantaged counterparts. Second, we disentangle the complex relationship between social status and rewards/costs for high achievement to show that score decrements for low status test-takers are affected by how socially and financially costly it will be to succeed. Third, we identify the mechanism by which disadvantaged students underperform: by way of a conscious decrease in effort. Fourth, we establish that there are particular conditions that exacerbate socially motivated underachievement; namely, when costs for success are implemented by an in-group member.

The studies that follow were designed to reflect the structural constraints that differentially affect high and low status individuals. We took great care in creating conditions that provide the strongest possible tests of our theoretically-derived hypotheses but that also mirror real-world counterparts. Complex social processes are, by nature, nuanced. We take advantage of the control afforded by experimental procedures to highlight the sequence by which status, cost framing, and group membership affect educational pursuits. While our contribution is to pinpoint the structural process that individuals in these situations experience, our expectation is that researchers with other methodological training and policy influence can use this research to similarly contribute

to the broader tapestry of understanding about how and why these social processes contribute to those of disadvantaged status underperforming on tests.

CHAPTER 2

BACKGROUND

The modern study of intelligence began with the idea that there is, at the same time, an underlying general ability that permeates all cognitive tasks and specific abilities that are unique to each cognitive task (Spearman 1927). The most widely accepted theory of intelligence to date, the Cattell-Horn-Carroll theory, builds on this idea, delineating intelligence into three hierarchically related stratum comprised of narrow, broad, and general abilities (Cattell 1941; Horn 1965; Carroll 1993). General ability (g) is an overarching cognitive ability that underlies a variety of other cognitive tasks. Fluid and crystallized ability (g-f and g-c, respectively), are further subsets of broad abilities. Fluid ability refers to the capacity to think logically, flexibly, and rapidly and to analyze and identify patterns in novel situations, whereas crystalized ability is more of a general collection of knowledge and vocabulary. These three abilities are the underlying constructs that most tests of intelligence aim to measure.

By definition, tests of intelligence measure the capacity to understand complex ideas, to adapt to the environment, to learn, to reason, to solve problems, and to overcome obstacles by thinking about them (Neisser et al. 1996). Researchers assess mental ability using psychometric measures, such as the Stanford-Binet, the Weschsler, and the Raven's Progressive Matrices (Terman and Merrill 1937; Wechsler 1939; Raven 1938). Each of these tests employs distinct items and procedures. However, both correlations between tests and factor analyses of subscales within tests indicate that all measure at least g, and

sometimes also g-f and/or g-c (Gottfredson 1997). Scores on intelligence tests are predictive of a range of related cognitive abilities and outcomes, such as school grades, job performance, and even the likelihood of engaging in criminal behavior (Gottfredson 2004; Fischer et al. 1996). Intelligence test scores are considerably stable throughout the lifespan with the largest deviations occurring in early childhood (Deary et al. 2000; Gow et al. 2011). While evidence does indicate some generational instability in intelligence scores, tests are re-normed approximately every ten years to account for this effect[1] (Flynn 2007). This indicates that such psychometrics are both reliable and valid measures of the construct of intelligence.

Despite that intelligence tests are some of the most psychometrically valid tests ever created, there are still major critiques about what exactly they measure and how differences in scores between groups highlight environmental influences on intelligence. Herrnstein and Murray (1994) assert that intelligence tests' major flaw is that they conflate social and natural predictors. In other words, these tests do not delineate the differences in intelligence that are caused by environmental factors versus genetic factors. This distinction is important because of the fear that those who originate from groups that have a lower average score will be seem as deterministicly low in intelligence, an argument that Herrnstein and Murray make amid controversy. Understanding the

[1] Dubbed the Flynn effect, raw intelligence test scores have increased by about 3 points every decade throughout the 20th century (Flynn 1987). This has been attributed to everything from industrialization to advances in mathematics instruction to more diversity in mating options to better nutrition (Daley et al. 2003; Mingroni 2004; Lynn 1990).

contribution of environment is important so as not to thwart efforts to intervene in environmental processes that can decrement scores (Fischer et al. 1996).

Biological Influences on Intelligence

Researchers have assessed that biological influences on intelligence both directly, through studies of the human genome, and indirectly, through heritability computations. To date, no single gene for intelligence has been isolated[2]. Rather, genetic markers of intelligence are widely distributed across many genes (Davies et al. 2011; Benyamin et al. 2014; Plomin et al. 2001). Still, some contributing genes have yet to be identified and the association of some seemingly known genes with intelligence have failed to replicate in other study populations (Chabris et al. 2012). At this point, too little is known about the complexities of the human genome for this area of research to contribute significantly to the nature versus nurture debate, other than to acknowledge that a portion of the variation in intelligence can be accounted for across a variety of genes.

While the direct study of the human genome is fairly recent, heritability studies have been around for decades. In fact, the major underpinning of the biological nature of intelligence stems from the fact that intelligence is one of the most heritable behavioral traits (Deary et al. 2009). Heritability is the ratio of genetic variation to total variation in a trait within a given population. Researchers usually test heritability using analyses of twins that have been reared together or apart. Based on studies of identical twins raised in separate environments, Bouchard et al. (1990) found that, despite different environments,

[2] It seems likely that no *single* gene will ever be identified. The dispersion of markers across many genes is consistent with other complex traits, such as height (Yang et al. 2010).

intelligence for separated twins was almost the same as that for non-separated twins. This shows that, irrespective of environmental influences, intelligence is strongly genetically related.

Such evidence is not as conclusive as it seems, however. Trait heritability, which is measured at the group level, is often misattributed to be equivalent to individual level genetic influence. In other words, it is not fair to say that since intelligence is highly heritable, the source of each individual's intelligence stems predominately from their genetic make-up. In fact, heritability coefficients for intelligence differ across races, socioeconomic statuses, and education levels. Advantaged groups, such as whites, those with high incomes, and the well-educated, have higher heritability rates than disadvantaged groups, such as blacks, those with low incomes, and the low-educated (Scarr-Salapatek 1971; Turkheimer et al. 2003; Rowe et al. 1999). This lower heritability means that there are environmental forces at large within these groups that do not allow group members to cultivate their genetic potential for intelligence (Nisbett et al. 2012). Furthermore, just as gene/environment interactions affect the expression of traits through epigenetics, highly heritable traits can still be highly modifiable (Sternberg 2012).

Environmental Influences on Intelligence[3]

Schiff et al. (1978) found that children adopted from deprived homes into affluent homes had higher intelligence than children who remained in deprived homes. In fact, the socioeconomic status of an adoptive family significantly affects the intelligence score of

[3] Chapter 4 goes into considerably more detail about the social processes that affect intelligence test scores and how they contribute to group differences in scores.

children (Capron and Duyme 1989; Duyme et al. 1999). This shows that there is something more than genetics that contributes to cognitive scores. The environment has significant effects on intelligence as well. In fact, research shows that poor pre-natal nutrition, exposure to lead, pre-natal exposure to alcohol, not being breast-fed, and stressful events experienced in-utero are all related to lower mental ability test scores (Pollitt et al. 1993; Needleman, Geiger and Frank 1985; Streisssguth, Barr and Sampson 1990; Mortensen et al. 2002; Eccleston 2011). These environmental influences are akin to suppressing the genetic potential of individuals.

Group Differences in Intelligence Test Scores

The two major test score gaps occur racially and socioeconomically. On average, low-income children score about 6 points lower than their high-income counterparts at age 2. This gap almost triples by age 16 (Von Stumm and Plumin 2015). Not only do children from disadvantaged family backgrounds have lower intelligence test scores than their advantaged counterparts (Bradley & Corwyn 2002; Schoon et al. 2012; Strenze 2007), but their test scores have been shown to worsen over time. This is in opposition to what happens to advantaged children whose scores are more likely to increase with time (Feinstein 2003). Along with test scores, socioeconomic status is strongly related to overall academic achievement (White 1982).

Beyond the obvious advantages that wealth bestows upon children, high-income families are also able to provide environments for their children that encourage intellectual growth. Children of professional parents heard 10 million more words than children of working class parents, providing them with a much richer vocabulary and an

obvious advantage on intelligence tests (Hart and Risley 1995). Lareau (2003) shows that parents of low socioeconomic status encourage obedience and respect for authority from their children, whereas wealthy parents teach their children skills that grow their intellectual curiosity. These attitudinal traits encourage upper-class children to pursue and enjoy the educational process. Hart and Riley (1995) quantified this, showing that the children of professional parents had a 6:1 encouragement to reprimand ratio, whereas children of working class parents had a 2:1 ratio. By way of different attitudes towards learning and schools, these patterns of socialization influence test scores, educational attainment, and future occupational success.

Racial differences in test scores are also prevalent, with score differences between blacks and whites often cited to be as large as 15 points (Neisser et al. 1996; Hernnstein and Murray 1994). Some research exists that shows that these differences are innate and due to genetic differences in behaviors, such as reaction times. However, such evidence is controversial. For example, most people that are labeled as black in America have a certain percentage of ancestry that would be labeled as white. Using this fact, Moore (1986) showed that the eventual intelligence test score of black and mixed race adoptees who had roughly equivalent intelligence test scores at age 8 was significantly affected by the race of the their adoptive family. Additionally, racial differences in test scores often do not take into consideration the institutionalized social policies that largely provide a white privilege (Fischer et al. 1996).

Racial minorities and low-income families are also much more likely to live in environments that have a negative effect on intelligence test scores. Black and low-income children are more likely to be exposed to lead and less likely to be breastfed

(Lanphear 2002; Forste et al. 2001; Mitra et al. 2004). Poverty-induced stress has been shown to lower intelligence test scores by 13 points on average by reducing the amount of mental bandwidth available to process information (Mani et al. 2013). Again, both racial minorities and the socioeconomically disadvantaged as more likely to experience these conditions, showcasing how these environmental discrepancies between groups can increase test score gaps.

While research supports the conclusion that both innate and environmental factors affect mental ability, individuals cannot significantly alter their genetic predispositions. Environments can be altered, however, and there is still much to learn about the effects of environmental factors on mental ability. Given that differences have been observed between individuals under differing social conditions, it is worth uncovering the mechanisms responsible for such findings.

CHAPTER 3

CONSEQUENCES OF INTELLIGENCE TEST SCORES

Those who use test scores to make decisions about the life of another person are much less concerned with how someone arrived at their score (through genetic giftedness or a superior environment) than they are with the actual score. In terms of practical implications, the theoretical concept of intelligence is replaced entirely by its measurable counterpart. Intelligence is the score that one receives on the test (Scarr 1997). And the consequences of the score are tremendous, touching all aspects of life, including educational attainment, occupational prestige, salary, and productive citizenship.

Intelligence Tests and Schooling

There is a positive association between intelligence test score and eventual level of education attained. Those with higher intelligence test scores tend to reach a higher level of education (Griliches and Mason 1972; Jencks et al. 1979; Hauser et al. 1983; Sewell et al. 1969). This relationship remains even after controlling for parent's level of education and other socioeconomic factors that are linked with educational attainment. One explanation of the higher educational attainment by those of high cognitive ability is the strong relationship between intelligence test score and academic achievement. Direct correlations tend to range from .40 to .50, with McGrew and Knopik (1993) even reporting a median correlation as high as .76. It is not surprising that those who do better at school stay in it for a longer amount of time.

In addition to test scores contributing significantly to eventual education level, school officials overtly use test scores to make decisions about educational settings for students. Administrators reserve placement into gifted and talented programs for students who have scores that exceed certain standards (usually above the 97th percentile of test-takers). Remedial services are also restricted based on test scores to students who have scores that are low enough to indicate a need for special education. School officials also use tests scores to track some students away from academics entirely, instead encouraging vocational courses. While the placement of students into these vocational categories is much more prevalent internationally, the practice does also occur domestically (Hanushek and Woessmann 2006). In all of these cases, access to school services are heavily, if not entirely, dependent on test scores such that these test scores quite literally determine the occupational path of the student.

Intelligence test scores also have major implications within the general student population. The practice of tracking, which occurs when schools group students into classrooms that are based on ability level. This type of sorting puts some students on a track where they receive a differentiated and enhanced educational experience from those who are tracked to a lower level. All students learn the same material, however, those with certain scores learn together. Students in higher ability tracks experience norms of engagement with the course material and high-quality teaching, all of which contribute to an ease of information retention and a culture of enthusiasm that is not present in lower tracks (Jennings et al. 2015; Barr and Dreeben 1983; Gamoran and Mare 1989; Lucas 2001; Oakes 1985).

The consequences of such tracking compound over the years. Imagine two students who are roughly equivalent, but separated by a point on an intelligence test. One student gets placed on the upper track and one on the lower. At first, this distinction seems arbitrary. However, by the end of schooling, the higher tracked student has actually outperformed the lower tracked student. Many factors are at play here. Higher tracked students are selected for their intelligence, then separated into isolated groups, and then given specialized training that is not available to other students. The placement of a student onto this higher track can be seen as a stamp of approval and confidence from school officials and can serve as somewhat of a self-fulfilling prophecy (Brophy 1983). Conversely so, for a lower tracked student.

If all else were equal, we might chalk this up to the existence of seemingly arbitrary, but procedurally necessary, cut-offs for school services. However, certain groups of people tend to end up on certain tracks. Low-income students are not nearly as prevalent in higher tracks as the wealthy (Gameran and Mare 1989). While, in theory, ability grouping and tracking allow schools to tailor educational experiences to their student population, in practice, they contribute heavily to educational inequality (Gamoran and Mare 1989; Lucas 1999; Oakes 1985). Scholars have likened tracking to re-segregation because of the additive effects that it can have on racial minorities (Zirkel and Cantor 2004).

Intelligence Tests and Occupations

Those who have higher intelligence test score tend to have more prestigious occupations. Gottfredson (1997) shows that the minimum threshold for an intelligence test score

increases with job prestige with the highest and lowest categories of occupations being separated by two standard deviations in median scores. This occurs through a number of different cognitively and environmentally-based routes. But even net of socioeconomic status and race, cognitive abilities remain integrally related to occupational prestige (Kerckhoff et al. 2001).

Hiring committees that select for these prestigious occupations may favor those of higher intelligence because intelligence predicts job performance in more or less all occupations (Hartigan and Wigdor 1989). In fact, as job complexity increases, the predictive validity of intelligence increases along with it (Hunter 1983; 1986). Those with higher intelligence could also be more in demand for jobs because they are easily trainable, owning to their ability to acquire skills independently and complete tasks without close supervision (Gottfredson 1997; Hunter and Schmidt 1996). Relatedly, cognitive ability is also associated with greater work productivity (Hauser et al. 2000)

As cognitive skills are strongly associated with both job performance and productivity, it follows that intelligent workers are highly rewarded by their employers (Jensen 1969). In fact, those with higher cognitive test scores tend to have much higher salaries than those with lower test scores. Again, group differences in occupational prestige and wages are prevalent. In fact, differences in intelligence tests scores explain a large portion (nearly half) of the racial wage gap (O'Neill 1990; Farkas and Vicknair 1996; Neal and Johnson 1996). Some attribute this difference, not to intelligence test scores alone, but to their effect on level of education received (Bowles and Gintis 2002).

Intelligence Tests and Community Contributions

Those who have higher test scores are more civically engaged. They tend to participate more in community-oriented activities, such as voting and taking part in rallies and other political demonstrations. They are more likely to sign petitions and to profess a greater interest in politics, in general (Deary 2008). In fact, Hernnstein and Murray (1994) go as far as to equate high intelligence with more civility and better citizenship. High scorers tend to use their money in a way that benefits others, such as investing. Intelligence test scores are also predictive of involvement in criminal behaviors with those of lower intelligence being more criminally active (Herrnstein and Murray 1994).

CHAPTER 4

SOCIAL PROCESSES THAT AFFECT TEST SCORES

Given the high stakes of test scores, it is crucial that test performance reflects an individual's actual abilities. However, previous research has shown that there are systematic differences in mental ability test scores that are not accounted for by differences in mental ability. Social processes, such as stereotype threat, socially motivated underachievement, and status-related bias alter an individual's test score away from their true score value, leading to decrements in scores for only certain categories of people (Steele and Aronson 1995; White et al. 2002; Lovaglia et al. 1998). These categories of people, namely, minorities and those of low-income, are disproportionally affected. Thus, understanding these processes could be key to lessening the achievement gaps discussed earlier.

Stereotype Threat

In general, test anxiety mediates the relationship between ability and test performance (Hembree 1988). Factors linked to test anxiety, such as stereotype threat, also can bias test performance so that it does not accurately reflect ability. Stereotype threat occurs in situations where there is pressure to avoid confirming a negative stereotype about one's group (Steele and Aronson 1995). For example, when told that people generally believe their group lacks high mental ability, black students performed worse on a test than when they were not primed with such a stereotype threat (Steele and Aronson 1995). Similar

effects have been found with those of lower socioeconomic status and disadvantaged racial categories (Croizet and Claire 1998; Desert et al. 2009; Aronson and Salinas 1997).

Stereotype threat not only impacts performance on tests, it also causes students to self-handicap in other ways, such as by decreasing their time spent practicing for tasks and by valuing the task, in general, less (Stone 2002; Aronson et al. 2002; Osborne 1995; Steele 1997). While any individual that fears confirming a negative stereotype about their group can be susceptible to stereotype threat, those who identify strongly with the stereotyped group tend to be the most affected (Ployhart et al. 2003; Schmader 2002). Explanations of why stereotype threat occurs tend to revolve around anxiety, but numerous other mechanisms have been noted, such as, but not limited to, reduced effort, physiological arousal, and a loss of motivation (Stone 2002; Osborne 2006; 2007; Schimel et al. 2004; Cadinu et al. 2005). Interventions have been successful in reducing instances of stereotype threat (Good et al. 2008; McGlone and Aronson 2006; Cohen et al. 1999). However, stereotype threat cannot account for all of the findings that our theory predicts, such as those involving non-stereotyped social characteristics.

Socially Motivated Underachievement

Along with test anxiety, performance has been known to suffer because of socially motivated underachievement. This occurs when individuals purposefully underachieve out of concern for others, to maintain relationships or to manage their social reputation (White et al. 2002). This differs from performance decrements due to social factors, such as stereotype threat, because individuals *consciously choose* to put forth less effort on academic endeavors. Socially motivated underachievers perceive high academic

performance in direct opposition to social acceptance (Fordham and Ogbu 1986, Fryer and Torelli 2004, Austin-Smith and Fryer 2005). If the norms of a social group discourage educational pursuits and social acceptance is a high priority for a student, then the academic performance of that student might not reflect their true intelligence.

The Burden of Acting White

The exploration of socially motivated underachievement started with Fordham and Ogbu's concept of the Burden of Acting White. This burden is activated when, to avoid being labeled as white, academically-capable black students purposefully underachieve. In the context of education, "acting white" means: spending time in the library studying, getting good grades in school, working hard at school, and speaking standard English (Fordham and Ogbu 1986). To avoid garnering this label, academically-capable black students might withholding efforts to achieve, opt out of selective courses and higher level tracks, and put more emphasis on sports-related talents than academic-related talents. This is particularly true when one's racial identity is high because the sanctions related to breaking social norms are much more costly to one who identifies more strongly with their group.

Empirical tests of the Burden of Acting White report contradictory evidence, however. While most studies can agree that blacks have experienced accusations of "acting white" from same-race peers (Datnow and Cooper 1997; Ferguson 2001), these accusations have not always been tied to academic performances (Bergin and Cook 2002). Additionally, black high achievers do not report lowering their educational aspirations or achievement because of accusations of "acting white" (Bergin and Cooks

2002; Datnow and Cooper 1997). Peer sanctioning for above-average performances are reportedly not widely prevalent and the social penalties that high achievers do receive are no greater for blacks than for whites (Bergin and Cooks 2002; Ainsworth-Darnell and Downey 1998; Cook and Ludwig 1997).

However, in support of Fordham and Ogbu's theory, Fryer and Torelli (2010) found that for blacks, as grades increase, popularity increases until a student reaches a high level of achievement (which they defined as a 3.5 grade point average). This relationship is particularly strong when considering data from a public rather than a private school and when the school's population is less than 20% black. Importantly, the relationship does not hold if students have a high socioeconomic status or if they attend private school. This shows the need for theoretical advancements, such as precise scope conditions, and methodological improvements, such as the use of experimental methods, in examining socially motivated underachievement.

Two-Audience Signaling Model

The Two-Audience Signaling model contributes significant theoretical advancements for socially motivated underachievement (Austen-Smith and Fryer 2005). Rather than concentrating solely on finding racialized performance decrements in school settings, it pinpoints the conditions under which underperformance might occur and those that are most susceptible to it. In this way, empirical tests can be more targeted towards the correct population.

The Two-Audience Signaling model proposes that blacks are pressured by two distinct audiences: the social audience and the economic audience. From the social

audience, they receive same-race peer pressure to conform to group norms, whereas, from the economic audience, they receive pressure to achieve academically so that they can ultimately increase their socioeconomic status. Individuals cannot succumb to the pressures of both audiences and are forced to choose which audience to impress. Essentially, they are deciding if they would prefer to be accepted by their peers or accepted to college.

The extent to which an individual values group acceptance relative to eventual income is the most important factor in determining who will succumb to socially motivated underachievement. Austen-Smith and Fryer (2005) present a typology of high and low social and economic types to showcase this. High social types, those that value group acceptance, fear peer group rejection more than low social types and are thus more susceptible to peer group norms of underinvesting in education. Low economic types expect to earn low wages regardless of their achievement. Rather than have both small wages and peer group rejection, low economic types purposefully underachieve as a method of ensuring their place amongst their group.

Experimental Tests of Socially Motivated Underachievement

In addition to the theoretical improvements described above, socially motivated underachievement has benefited from numerous experimental tests in the domain of psychological social psychology. Studies show that high performers experience emotional distress and are fearful of rejection and group sanctions (Exline and Lobel 1999; 2001). In reaction to knowingly outperforming others, participants avoid those that they outperformed and/or underperform themselves in order to fit in with group norms

(Exline et al. 2013; White et al. 2002). A fear of outperforming others also encourages students to prefer private, rather than public, recognition of their achievements (Exline et al. 2004). While these studies make only passing allusions to applications in racially and socio-economically driven socially motivated underachievement, they point to a very similar process.

Status and Test Scores

Most relevantly to the study at hand, Lovaglia et al. (1998) demonstrated that social status impacts mental ability test scores. Within a given culture, those who possess qualities that are considered to be low status (e.g., racial and ethnic minorities) do worse on tests than those who possess high status qualities. This stems from a process by which those of low status expected lower rewards and higher costs for high achievement than did their high status counterparts.

These expected consequences are paramount to understanding status-related score decrements. For example, stereotype threat asserts that students expect to receive sanctions for scoring low on a test. This low score would, in turn, confirm negative stereotypes about their group. Status-related bias, on the other hand, affects those who expect sanctions for scoring too high. This high score would signal norm violations to their group (Lovaglia et al. 2004). It may be the case, however, that costs for scoring too low via stereotype threat are perceived as costs from an out-group, whereas costs for scoring too high via status-related bias are viewed as coming from in-group members. In this way, the source of the sanctions might help to discriminate between stereotype threat and status-related score decrements.

Building on these studies concerning status-related bias, we draw on Status Characteristics Theory (SCT) for our arguments concerning mental ability testing. While SCT is typically concerned with collective tasks involving two or more interactants, an extension of SCT states that individual performance will also be a function of status processes if individual performance is used to determine future status (Lovaglia and Lucas 1997). Although Lovaglia et al. (1998) demonstrated that status processes affect test scores, there have been few follow-up studies for almost two decades.

CHAPTER 5

THEORY

Status Characteristics Theory

Status Characteristics Theory (SCT) links the social roles and expectations from larger society to stratification processes in small groups. Through a series of logically interrelated assumptions, precise term definitions, and scope conditions, SCT connects these culturally specific beliefs to group members' expectations of task performance for both themselves and others. These expectations, in turn, influence the actual performance of the members of the group. In this way, SCT makes predictions about how members of groups will become stratified and who will have the most influence over group decisions. SCT is part of a larger body of work on expectation states, with research evidence that dates back approximately 55 years (Berger 1958).

SCT rests on the idea that there are two kinds of characteristics that group members find integral to tasks. Specific status characteristics are characteristics that create expectations for performance in limited settings. Examples include specific abilities like basketball skills or mathematical expertise. If a group gets together to solve a calculus problem, it would make sense that they might base their expectations for performance on a specific status characteristic, such as mathematical ability. However, group members also base their opinions of others on characteristics that are not actually related to the task. Such characteristics, called diffuse status characteristics, create expectations for performance in diverse situations. Examples include race, gender,

education level, physical attractiveness, and age (Cohen and Roper 1972; Webster and Driskell 1978; Foschi 1996; Pugh and Wahrman 1983; Wagner, Ford, and Ford 1986; Walker et al. 1996). Both the specific and diffuse status characteristics that each group member possesses aggregate to form overall performance expectations that contribute to a group member's place in the group's status hierarchy. These hierarchies are unconsciously formed, collectively agreed upon, and result in high and low status group members (Berger et al. 1977). Costs and benefits are bestowed upon group members according to their relative status in this hierarchy. Those of high status tend to participate more in groups, be more positively evaluated by other group members, are more likely to be elected as group leaders, and have, overall, more influence of the group's eventual decisions. Those of low status participate less in group tasks, have their suggestions more negatively evaluated by other group members and tend to have less influence over their fellow group members (Berger 1958; Berger and Connor 1974).

The predictions that SCT makes apply only in particular situations. The following scope conditions outline such instances (Berger et al. 1977). There must be at least two individuals working together on a task that can be evaluated in terms of success or failure. The individuals must be motivated towards successful task completion and believe that there are some individuals that possess a characteristic that is instrumental to completing the task, while there are others who do not. The task must also be completed collectively amongst group members. Many groups fall within these scope conditions, such as hiring committees, juries, and classroom workgroups.

The earliest research on the impact of status characteristics showed that both high status and low status actors evaluate the contributions of high status actors more

positively than the contributions of low status actors (Berger et al. 1977). Subsequent research further showed that status effects persist across interaction settings (Pugh and Wahrman 1983; Markovsky et al 1984). From SCT and its associated Reward Expectations Theory (Berger et al. 1998), we also know that status processes generate social structures wherein rewards are inferred based upon status. However, we suggest that low status actors also may anticipate costs associated with high performance if such performance would conflict with the status structure. Assuming no actual status-based differences in ability, if low status actors anticipate higher costs for high performance, then it is in their best interest to underachieve. In this way, performing lower than their actual ability spares them from receiving penalties.

In order to apply SCT in this way, there must be shared expectations that low status actors have lower ability and, more than for high status actors, that high performance will be costly for those lower in status. The anticipated consequences of high test performance thus explain the differences in test scores amongst high and low status actors. These consequences include expectations on the part of low status actors that they will be penalized for high performance, with no such expectation held by high status actors (Lovaglia et al. 1998). SCT treats different status characteristics as having identical consequences for interaction. Characteristics that have been shown to produce expectations for competence, regardless of actual competence, include age, race, gender, education level and physical attractiveness. Whether treating African-Americans as lower status than Caucasians (Cohen and Roper 1972; Webster and Driskell 1978), or females as lower status than males (Foschi 1996; Pugh and Wahrman 1983; Wagner, Ford, and Ford 1986; Walker et al. 1996), the higher the "state" of the characteristic one possesses,

the more influence one is assumed to have in task-relevant interactions. Therefore, if low status actors tend to do worse on mental ability tests than high status actors, this effect should present itself both in terms of race in mixed-race settings, and gender in mixed-gender interactions.

The mental ability test effect does not occur in mixed-gender interactions, however. On average, females do as well as males on such tests (Jensen 1998; Stumpf and Stanley 1996). Still, differences in scores do appear along racial lines. On average, African-Americans do worse than Caucasians on such tests (Jensen 1998; Herrnstein and Murray 1994). For these reasons, there must be a mechanism other than status that affects test performance. Earlier we conflated rewards and costs when discussing anticipated consequences of mental ability test performance. As we discuss below, however, there is strong evidence that people do not treat rewards and costs symmetrically: the motivation to avoid costs is stronger than the motivation to seek comparable rewards. We propose that one reason intelligence test scores differ across racial lines but not across gender lines is that females expect lower rewards than males for scoring high on mental ability tests, while African-Americans expect higher costs than Caucasians for scoring high on mental ability tests. In this way, the perception of rewards and costs in low status actors is one of the mechanisms responsible for lower test scores.

Rational Choice Theory and Prospect Theory

Rational Choice Theory (RCT) holds that individuals assess their options and make choices based on the expected benefits and costs of these decisions (Homans 1961). RCT makes the follow assumptions about individual behavior. All choices can be rank ordered

in terms of preference (completeness) and all choices can be compared to other choices (transitivity). Additionally, alternative choices that are not within the choice set would not encourage a re-ordering of the choices that are within the current choice set (independence of irrelevant alternatives). While individuals do seek to maximize the utility of their choices, they only act as if they are aware of and can calculate the benefits and costs of each alternative.

Within this framework actors make decisions based on rewards and costs, preferring large rewards to small rewards, and small costs to large costs (Simon 1955). Status characteristics also have been shown to generate reward expectations consistent with status levels (e.g., Harrod 1980). It follows that, because rewards and costs are strongly correlated with status, those with low status expect smaller rewards and/or larger costs. However, as stated above, perceptions of rewards and costs may differ across status characteristics. Actors in the low state of some status characteristics may associate smaller rewards with high performance on mental ability tests, while actors in the low state of other status characteristics may associate larger costs with high performance on the same tests.

SCT assumes that states of status characteristics receive their differential evaluations from the broader culture within which the focal group interaction transpires. It thus seems plausible to consider the possibility that the broader culture may also emphasize the advantages rather than the disadvantages of certain status characteristics, and the disadvantages rather than the advantages of others. Our purpose in this research, however, is not to identify the source of these distinctions if they do in fact exist, but rather to test for their effects.

According to Prospect Theory (PT), these differing perceptions would affect choice behavior: all else equal, actors are more likely to engage in behavior to avoid costs than to acquire rewards (Kahneman and Tversky 1979; Tversky and Kahneman 1992). PT, which was developed as an alternative model to RCT, asserts that there are two stages that individuals use to evaluate choices. Similar to RCT, options are first ordered in relation to preference. However, individuals then act as if they are setting a reference point with which to compare all of the options. In relation to this reference point, anything higher is seen as a benefit, whereas anything lower is viewed as a cost. In this way, PT differs from RCT in that it is more concerned with relative, rather than absolute, gains and losses. This sensitivity to potential losses may explain the difference between the test scores of low status actors that perceive high costs associated with high performance vs. low status actors that perceive lower rewards associated with high performance.

Self-categorization Theory

The in-group/out-group status of the source of these rewards and costs may be equally as important to this process. Just as individuals do not perceive rewards and costs symmetrically, they do not regard the source of these rewards and costs as equal either. A cost for high achievement may loom larger, if it comes from an in-group member. Self-categorization theory (ST) details how this process occurs (Turner 1985; Turner et al. 1987). Individuals view their connections to groups in terms of shared social categories. These categories become pervasive to their own identity such that they experience a process of depersonalization. Depersonalization occurs when group members perceive

themselves as interchangeable members of the group that possess prototypical features of the group (Turner 1985). Such prototypically involves agreement with and portrayal of the beliefs, opinions, goals, and norms of the group. Consensus with in-group members generates positive feelings, whereas disagreement with in-group members generates uncertainty (David and Turner 1996, 2001). Identity processes other than ST have also been noted to contribute to score decrements. When student that they are taking a test that confirms their identity, they do much better than when they think that success on the test will confirm a contrasting identity (Youngreen et al. 2004).

The Burden of Acting White, discussed earlier, provides an excellent explanatory vehicle for how achievement is affected by identity processes (Fordham and Ogbu 1986). Using this example, we assume that a group norm is to dis-identify from educational pursuits. If a black student engages in a behavior that violates the norms of a group, such as participating frequently in class or always completing homework assignments, then they are acting in direct opposition to their group. Rather than blending into the group via depersonalization and prototypicality, they are instead marking themselves as an outsider. In this example, the cost of violating the group norm is essentially ostracism from the group with the label of "acting white". Depending on how much a student identifies with their group, this cost could be enough to deter them from high scholastic achievement.

Formal Theory

According to PT, actors should be more affected by the threat of costs than to the chance of rewards. ST implies that individuals will also be more sensitive to costs, if they

originate from their in-group, rather than their out-group. This leads to Propositions 1 and 2:

P1: All else being equal, an actor will seek to avoid costs rather than obtain rewards.
P2: All else being equal, an actor will seek to avoid costs from their in-group more than costs from their out-group.

Lovaglia et al. (1998) found that low status actors scored lower than high status actors on a measure of mental ability. However rewards and costs were confounded in that research. Our research teases apart their effects by determining whether low status actors respond differently to the prospect of gaining smaller rewards for high scores, or incurring higher costs for high scores. The following sections use gender and race to illustrate this argument in more concrete terms.

Low Rewards for High Performance

Traditionally, females can expect lower rewards for high performance than males. When females hold high status positions such as lawyer, professor, engineer, physician or business executive, their average salary is significantly less than the average male's salary (Katz and Murphy 1992; Blau and Kahn 1994; Wood et. al 1993; Biddle and Hamermesh 1998; Barbezat 1987; Gander 1997; Morgan 1998; Bertran and Hallock 2001). Similarly, although women receive equal opportunities for promotion, they are held to higher standards of performance (Olson and Becker 1983). Furthermore, when females are promoted, they receive a smaller salary increase than males (Gerhart and

Milkovich 1989) and are accorded less prestige (Tyree and Hicks 1988). This leads to Proposition 2a:

P2a: Lower status actors will expect lower rewards for high performance than higher status actors.

High Costs Associated with High Performance

Traditionally, African-Americans can expect higher costs for high performance than Caucasians. Research shows that minority students who do well on tests can expect to be socially shunned by their peers for trying to "act white" (Fordham and Ogbu 1986; Steinberg, Dornbusch and Brown 1992). African-Americans that portray academic achievement in a positive light are more likely to be depressed and concerned about social approval (Arroyo and Zigler 1995). Additionally, African-Americans who attempt to complete a college degree report high personal and emotional costs. The social costs may be so high that they come to expect that the degree will not improve their lives (Steele 1992).

African-Americans can also expect to be penalized for their mental ability if it exceeds the expectations of teachers. This "Pygmalion Effect" occurs when African-Americans with high IQ scores are criticized more and praised less than both Caucasians with high and low IQ scores and African- Americans with low IQ scores (Rosenthal and Jacobson 1968; Rubovites and Maehr 1973). For these reasons, African-Americans can expect to incur costs for scoring higher than expected on mental ability tests from both

their in-group (their peers) and their out-group (their teachers). This leads to the following propositions:

P2b: Lower status actors will expect higher costs for high performance than higher status actors.

P3a: Low status actors will expect high costs from in-group members for high performance.

P3b: Low status actors will expect high costs from out-group members for high performance.

Low Rewards vs. High Costs

If the salient message to African-Americans is "keep your head down and do not stand out" and the salient message to females is "work twice as hard to get half as far," then this would be reflected in the perceptions of rewards and costs associated with high performance. For this reason, we will test for the effects of rewards and costs on low status actors as related to their scores on mental ability tests. The following derivation (D) obtains from the foregoing propositions:

D1: All else being equal, if an actor is low in status and expects lower rewards and higher costs for high performance, s/he will seek to avoid costs rather than to obtain rewards.

In-Group Costs vs. Out-Group Costs

If a group member identifies with their group, then they will place more emphasis on the opinions of fellow group members than outsiders. In-group sanctioning for high achievement, therefore, would serve as more of a deterrent for high performance than out-group sanctioning. We derive the following from the foregoing theoretical discussion on group identification processes:

D2: All else being equal, if an actor expects high costs from an in-group member, they will seek to avoid these costs more than if these costs come from an out-group member.

Thus, in addition to potentially replicating Lovaglia and colleagues' important findings, our research teases out the potentially separate effects that perceived social rewards and social costs may have on intelligence test performance and makes predictions about when this process is more likely to occur; that is, in the presence of in-group sanctions.

CHAPTER 6

STUDY ONE: METHODS

We instantiated the above theoretical derivations and conducted two experiments that were designed to isolate the effects of status, costs/rewards for high achievement, and group membership on test scores. As detailed below, Study One examines how status and cost/reward framing intersect to cause score decrements in test-takers. Study Two extends Study One by looking at the relationship between in-group/out-group memberships of the person who makes achievement costly and test scores. Throughout the design of these studies, we relied on well-tried methods and measures. However, when necessary, we pre-tested novel experimental manipulations and different versions of the dependent variable. Appendix A fully details these tests.

General Procedures

Upon arrival at the lab, each subject was met by a project assistant and escorted to a small room with a desk, keyboard and computer monitor. Before proceeding, the subject was asked to read and sign an informed consent form. The assistant then initiated a program on the subject's computer[4]. After a brief introduction to the study, the computer

[4] Authorware was used to program the subject interface, present information and gather responses. Along with programming done b the authors, we used programs and scripts designed by Will Kalkhoff and Michael Lovaglia.

administered a questionnaire requesting information on age, gender, college and high school GPA, SAT scores, mother's and father's occupation and education history, and an estimate of how many books were in their house as a child. This was also used by Lovaglia et al. (1998), with the measured factors serving as potential control variables in statistical analyses. Following the collection of demographic and control variable information, subjects were told that they would take a test that would influence which pay grade would determine their compensation for the study. They were told that their pay would be based on both their ostensible status as right or left-handed (see below) and their performance on the test. All the while, instructions and pay grades were presented with language that sounded rewarding or costly. This constituted the framing manipulation.

After completing the test, each subject was interviewed by a trained lab assistant. The purpose of the interview was to ensure that the subject was fully engaged in the study, understood all instructions, and had no suspicions with respect to manipulations or deceptions. Following the interview, the subject was debriefed as to the overall purpose of the research, the hypotheses being tested, and his or her experimental condition. The nature and purpose of the deceptions were also explained in the debriefing, and the assistant offered repeated opportunities to ask questions and seek clarifications. The assistant asked the subject for a verbal commitment to not share any details of the research with other students, as this would eliminate others' chances of participating. Before leaving, the subject was paid in cash for his or her time, thanked by the assistant, and escorted to the exit door.

Design

We used a 2 (Status: High, Low) x 2 (Framing: Rewards, Costs) factorial design. The Status factor accounted for whether participants were made to believe that their handedness would either help (High Status) or hinder (Low Status) their effort at the group task, while the Framing factor was based on whether pay for the study was framed in terms of gains (Reward) or losses (Cost).

Independent Variable: Status Manipulation

Rather than relying on naturally occurring status differences among subjects, we capitalized on the laboratory setting by controlling perceived status differences. When testing for status effects, we can differentiate subjects by one clearly defined task-relevant status characteristic. This experiment employed a status manipulation phase in which participants believed they possessed either the high or low state of a status characteristic. This was accomplished as follows. The computer program asked the subjects to identify as either right or left-handed (in reality, the study was given exclusively to right-handed subjects for the simplicity of programming and initiating only one subject interface). Subjects were then informed that in the upcoming half of the study they will be working on a difficult task requiring group members to work together in order to complete it successfully. They were told that ability at this task is strongly related to handedness. In fact, there was no group interaction as we were only interested in the impact of the status assignment on performance at the mental ability test to follow. Using the same method as Lovaglia et al. (1998), we provided several rationales to associate either the subjects' non-dominant hand with traits related to the successful

completion of the task, and his/her dominant hand with traits related to unsuccessful task outcomes or vice versa depending on the status condition (see Appendix B). For example, to make a participant think that left-handers were higher in status, they would be told that left-handers were much better at coordinating group work, getting things done in groups, and leading groups. They would also be told that right-handers, in contrast, were too inflexible, made poor decisions, and were so rigid that they could not lead a group effectively. Therefore, based on handedness, subjects were either assigned to the low status or high status position. Methods of establishing status based on seemingly arbitrary characteristics have been used successfully in much prior research (Jemmott and Gonzalez 1989). Mild deceptions such as these are extremely common in experimental research on status. In every case, they are identified in a debriefing at the end of the study and their purpose explained thoroughly to subjects.

Independent Variable: Framing Factor

Subjects were informed of a series of pay categories associated with performance on a fictional work aptitude test. They were given information showing salaries for the ostensible upcoming group task that ranged from $10-$38, and whether a low or high status actor (in terms of handedness) could obtain this salary.

Table 6.1. Reward Framing Manipulation.

Status	Aptitude	*Job*	*Pay*	Status	Aptitude
Left-handed	Superior	***Psychotherapist***	***$38***		
Left-handed	Moderate	***Technician***	***$13.50***	Right-handed	Superior
Left-handed	Poor	***Orderly***	***$10.00***	Right-handed	Moderate, Poor

Table 6.1 shows the framing manipulation for a low status subject whose study was framed in terms of rewards. As the table shows, rewards for doing well on the test were not as high as if that subject were of high status. It is impossible for them to obtain a job paying higher than $13.50. However, a high status subject could earn up to $38.00. All subjects were also told that high scoring high status subjects would receive recognition from the career center for their outstanding performance and that this kind of recognition is useful for obtaining jobs after college. Low status high scoring subjects, on the other hand, would simply receive a pamphlet detailing the resources available at the career center. These manipulations were intended to reflect the different outcomes for high and low status individuals in terms of high and low rewards.

Table 6.2. Cost Framing Manipulation.

Status	Aptitude	*Job*	*Pay Cut*	*Pay*	Status	Aptitude
Left-handed	Superior	***Psychotherapist***	***-$0***	***$38***		
Left-handed	Moderate	***Technician***	***-$24.50***	***$13.50***	Right-handed	Superior
Left-handed	Poor	***Orderly***	***-$28.00***	***$10.00***	Right-handed	Moderate, Poor

Table 6.2 outlines the framing manipulation chart for a high status subject whose study was framed in terms of costs. Subjects were told that they would typically earn up to $38 for their participation in the second half of the study. However, reflecting difficult economic times, the organization would be cutting salaries. As in the Reward conditions, they were told that their salary would be determined by their handedness and by their aptitude test score. However, emphasis was placed on the fact that this was a salary cut, rather than just a payment to them. Salaries were adjusted down from $38, with cuts ranging from $0 to $28.00 for high status participant. Low status subjects were guaranteed a salary cut ranging from $24.50 - $28. For low status subjects, costs for doing well on the test were larger than those of ostensible high status others. They could never receive a salary cut less than $24.50. However, a high score on the aptitude test would still result in a smaller salary cut than if they scored lower on the test. Thus, low status subjects are incentivized to obtain mediocre scores; enough to get by, but not extraordinarily high. In addition to the financial costs associated with performance, subjects were also told that since low status subjects tend not to score very high on these kinds of tests, when they do, it is usually due to cheating. High status subjects did not face such accusations.

Mediating Variable: Effort

Remember that a major focal point of socially motivated underachievement is that students consciously choose to underperform. To assess this, we used both a self-report and a behavioral measure. Immediately following the ostensible work group test, participants rated statements that evaluated their effort, the importance of the test to them,

and their overall motivation to do well (see Appendix C). This measure has been used in much prior research into motivation and test-taking (Sundre and Moore 2002). We also measured the time that students spent taking the test to approximately a tenth of a second. We measured this at the test level, rather than at the item level, in order to obtain a measure of the effort given overall. Time spent taking a test has been linked to effort in previous work (Wise and Kong 2005).

Dependent Variable: Mental Ability Test Score

We assessed mental ability using the Advanced Raven Progressive Matrices test (see Appendix D for a sample question). The "Raven" provides reliable and valid measures of mental ability and has been shown to correlate well with other measures of general intelligence (Raven, Court and Raven 1992; Jensen 1992). We chose it because it does not contain questions that are typically recognized as mental ability test questions. This reduces the chances that subjects will be affected by a priori performance expectations. Additionally, because there is no reading involved in the test, it is less culture-bound and thus less biased against minorities (Rushton et al. 2004). Furthermore, the Raven is self-administered, which eliminates the chance for any bias by the test-givers. Based on a pre-test of two versions of the Raven of varying difficulty, we found that the Advanced version was more suitable for a college population (see Appendix A). The use of the test is rationalized to subjects by informing them that it will determine their role in the ostensive second part of the study.

Hypotheses

This research design teases apart the effects that those of differing status experience due to differing reward and costs structures. The following two hypotheses confirm previous work on status and cost framing (Lovaglia et al. 1998; Tversky and Kahneman 1992):

H1: There will be a main effect of Status. Those who are made to feel of low status will score lower on the Raven's than those who are made to feel of high status.

H2: There will be a main effect of Framing. Those who perceive their study in terms of Costs will score lower on the Raven's than those who perceive their study in terms of Rewards.

Uniquely, we also expect status and framing to interact to produce the largest score decrements of all. Earlier we derived that all else being equal, if an actor is low in status and expects lower rewards and higher costs for high performance, s/he will seek to avoid costs rather than to obtain rewards (D1). By substituting our operational instantiations of the foregoing theoretical discussion, we hypothesize the following:

H3: There will be an interaction effect of Status and Framing. Those who are made to feel low status and who perceive their study in terms of Costs will score lower on the Raven's than all else.

CHAPTER 7

STUDY ONE: ANALYSES

Subjects

Subjects were recruited from undergraduate classes at the University of South Carolina and assigned randomly to experimental conditions. All subjects were prescreened to be white, right-handed females. Approximately 38 subjects participated in each condition, a number that in comparable studies has proven sufficiently powerful to detect moderate differences across conditions[5]. Each subject was paid $10 for participation in the study.

Data

Data was excluded for one of the following reasons: 1) lack of task orientation or comprehension (3 participants); and 2) disbelief that handedness contributed to an increased/decreased task ability (16 participants). This left 152 analyzable cases[6].

Of these cases all were female. All reported their race as white in the pre-study, however during the actual study, five people categorized themselves as White-Hispanic or Other. We kept these cases in the analysis because of their previous reports and because this should not theoretically effect results.

[5] See Appendix E for an analysis of statistical power to detect differences between conditions.

[6] Studies using the standardized experimental setting typically eliminate about 15% of cases due to scope violations. Our exclusion rate of 12.5% is typical for SCT studies (Dippong 2012).

Regression Analysis

Regression Diagnostics.

Prior to analysis, data were checked to ensure that they upheld the assumptions of linear regression. Data were normally distributed, homoschedastic and no collinearity was present. Prior to model building, we ran an analysis for multivariate outliers on the full model using Cook's Distance. We eliminated 8 cases because they had a distance greater than 4/n (Bollen and Jackman 1990). This means that these cases have high leverage and are highly influential. Thus, they can bias the regression analysis by showing undue weight.

Model Building.

We began our analysis by including the main effects, interaction and all of the control variables in the model. However, an F-test for nested models showed that constraining High School GPA and College GPA resulted in improved model fit, $F(2,137)= .184$, $p= .832$ (see Appendix F for calculations). We chose use this more parsimonious and better fitting model, despite that the results were robust to the inclusion of these variables because keeping these two variables in the model would be equivalent to parameterizing noise (see Table 7.1).

Final Model.

Controlling for college admission test scores, we found a significant interaction effect of status and framing. Low status subjects scored the worst on the Raven when they believed that a high performance would be costly to them, $b= 2.927(1.304)$, $p< .05$. This

confirms Hypothesis 1. Status and framing also each have an independent main effect on test scores, b= -2.326(.894), p< .01 and b=-2.647(.900), p< .01, respectively, thus confirming Hypotheses 2 and 3 (see Table 7.1).

Those who felt low status scored almost a full point lower on average than those who felt high status. Subjects who believed that their scores would be associated with high social and financial costs scored about 1.5 points lower on average than those who believed that their scores would be associated with low social and financial rewards (see Table 7.2). The largest score decrement occurred when low status subjects were put in a position in which they believed that they would endure costs for achievement. Their scores were lowest of all, nearly 2.5 points below the highest scores and over a full point below the average test-taker.

Table 7.1. F-Test of Nested Models: Study One.

	Model 1 B (SE)	*Model 2* B(SE)
Intercept	25.480 (4.401)***	25.384 (.633)***
Independent Variables		
Status	-2.323 (.900)*	-2.326 (.894)**
Framing	-2.622 (.914)**	-2.647 (.900)**
Status*Framing	2.876 (1.324)*	2.927 (1.304)*
College Admission Test Scores	2.360 (.355)***	2.374 (.338)***
College GPA	1.892 (3.918)	
High School GPA	-2.029 (4.122)	
Omnibus F	9.545	14.394

*Note: *=p<.05, **=p<.01, ***=p<.001.*

Table 7.2. Predicted Marginal Means.

Manipulation	*Raven Test Score*
High Status, Reward Frame	25.111 (2.325)
High Status, Cost Frame	23.222 (2.079)
Low Status, Reward Frame	23.842 (2.091)
Low Status, Cost Frame	22.647 (2.500)

Mediation Analysis.

In addition to the main effects and interaction, we tested a mediation model to determine why low status subjects whose study was framed as costly tended to have lower scores than everyone else. The mediator in question was a behavioral indicator of effort measured by the amount of time spent on the test. We are confident that time spent on the test was an accurate measure of effort, rather than some other indicator because time spent on the test was highly correlated with self-report measures of effort, but not with self-report measures of test importance (see Table 7.3).

Table 7.3. Pearson r Bivariate Correlations.

	Self-Reported Effort	Self-Reported Test Importance
Time Spent on Test	.228**	.020

*Note: **=p<.01*

We first determined that independent variables significantly predict the dependent variable. As evidenced above in Table 7.1, controlling for college admission scores, status, framing, the interaction of status and framing all significantly predicted Ravens test score. Table 7.4 identifies that only status predicts the mediator variable of effort, b= -5.836 (2.133), p< .01. This means that we can only continue testing whether the status effect is mediated by effort. Table 7.5 shows that the effort is significantly related to

Ravens score, b= .168 (.033), p <.001. Finally, when effort is accounted for in the model, status loses its effect on Ravens score, b= -1.343 (.844), p=ns. Additionally, Sobel's Test shows that the indirect effect is significant, -2.410 (.407), p<.05. Given these calculations, effort significantly mediates the relationship between status and test score. This means that low status individuals scored lower on the Ravens test because they put forth less effort.

Table 7.4. Independent Variables Predicts Effort.

	B(SE)
Intercept	32.814 (1.509)***
Independent Variables	
Status	-5.836 (2.133)**
Framing	-2.184 (2.146)
Status*Framing	4.991 (3.110)
Standardized Test Scores	2.172 (.806)**

*Note: *=p<.05, **=p<.01, ***=p<.001.*

Table 7.5. Effort Mediates the Effect of Status on Raven's Score.

	B(SE)
Intercept	19.858 (1.220)***
Independent Variables	
Status	-1.343 (.844)
Framing	-2.279 (.830)**
Status*Framing	2.087 (1.209)
Standardized Test Scores	2.009 (.319)***
Time Spent on Test	.168 (.033)***

*Note: *=p<.05, **=p<.01, ***=p<.001.*

CHAPTER 8

STUDY TWO: METHODS

General Procedures

The procedures of Study Two largely mirrored that of the first study. Participants were guided through a computer program that collected demographic information, presented the experimental manipulations, and administered a test that would ostensibly place the subject in a job and pay grade for the second part of the study. Again, we were only interested in the effects of the experimental and control variables on the test score and, as such, no group work ever occurred.

Design

We used a completely randomized design wherein all subjects were made to believe that they were of low status and that success on the test would be both socially and financially costly. Conditions differed based on the expected source of the costs. Either an in-group member or an out-group member was thought to be responsible for creating the costly conditions that discouraged success on the test.

Invoking Group Membership

Participants believed that both students from the University of South Carolina and its rival institution, Clemson University, were enrolled in the study. Following demographic questions, their identity as a student at USC was invoked by having students rate how

much they like being a student at USC and by writing why they chose to attend the USC over Clemson University. They were then told that

> *According to research, people who attend the same university are likely to be more similar to one another than they might expect. These similarities include several dimensions, such as worldview, values, attitudes and behaviors. This means that those who also attend the University of South Carolina may be more similar to you than those who attend Clemson University. We do not know all of the reasons why those who attend the same college tend to be more similar to one another than those who attend different colleges, but we do know that this similarity exists and has been shown in many other research studies. One of the goals of today's study is to find out more about this.*

We assessed the strength of participant's identification with the in-group by asking them a series of questions that were designed to evaluate their identity level. Questions included how much participants saw themselves as group members, how much they felt strong ties to other group members… etc. (see Appendix G). We collected responses to these questions both before and after the status manipulation to ensure that it did not negatively affect a participant's willingness to identify with the group.

Independent Variables: Status Manipulation

Again, we controlled perceived status differences. However, this time, participants believed that the members of a group with which they were affiliated, their university, possessed either the high or low state of a status characteristic. After having the subject select if they were affiliated with the University of South Carolina or Clemson University, the computer program informed them that in the second half of the study they will be working on a difficult group task that requires cooperation. They were told that ability at this task is strongly related to institutional affiliation. We provided several justifications to associate members of the subjects' rival college affiliation with traits that are integral to the successful completion of the group task. At the same time, we associated members of the subjects' institution with traits related to unsuccessful task outcomes (see Appendix H). Therefore, all participants were assigned to the low status position based on their institutional affiliation. As we showed in Study One and in the pre-test sections (see Appendix A), these methods are quite successful at turning somewhat arbitrary characteristics into markers of status distinction. Participants were informed of all deceptions during the debriefing session that followed the study.

Independent Variables: Framing Manipulation

We used the same framing manipulation as that of Study One, except that all subjects experienced costs associated with success. Table 8.1 outlines the framing manipulation chart. Remember that all subjects were made to think that those who attend the University of South Carolina do worse at group tasks than those who attend Clemson University. In this way, costs for doing well on the test were larger for those of low status

than those of ostensible high status. Subjects were also told that the high scores of low status participants were suspect because of the likelihood that they were due to cheating. The treat of a referral to the academic integrity team was meant to invoke a social cost.

Table 8.1. Cost Framing Manipulation.

Status and Group	**Aptitude**	***Job***	***Pay Cut***	***Pay***	**Status and Group**	**Aptitude**
USC	Superior	***Psychotherapist***	***-$0***	***$38***		
USC	Moderate	***Technician***	***-$24.50***	***$13.50***	Clemson	Superior
USC	Poor	***Orderly***	***-$28.00***	***$10.00***	Clemson	Moderate, Poor

Independent Variable: Source of Costs for High Achievement

Since participants thought that both students from the University of South Carolina and its rival institution, Clemson University, were enrolled in the study, this allowed us to convince them that either an in-group member (a student at the University of South Carolina) or an out-group member (a student at Clemson University) was responsible for creating the job and pay charts that were associated with the ostensible second part of the study. They were told that

> *To try to optimize group performance, we ran a previous study to determine the best way to pay you for your participation in this study. In this previous study, we asked [students] to rate pay charts. We asked them to select the chart that they thought should be used to calculate your pay for the study. University of South Carolina (Clemson University) students*

decided that your pay should be determined by your college affiliation and by your score on the aptitude test.

Participants then viewed the framing manipulation chart detailed above. Now, they not only face costs for success, but they know who was responsible for setting up this unfair system, either their in-group or their out-group. They were also told that ordinarily everyone gets paid more for their participation in the study, but that either their in-group or out-group had decided that the tough economic times warranted a decrease in pay for everyone else. In this way, social and financial costs for high achievement were either imposed by an in-group or out-group member. Manipulations such as these have been used in much prior research and have been successful at eliciting group-minded behavior (Tajfel 1979).

Hypotheses

This research delineates how in-group and out-group sanctioning affects test scores. Theoretically, all else being equal, if an actor expects high costs from an in-group member, they will seek to avoid these costs (D2). Again, we hypothesize the following by substituting measurable instances of the theoretical arguments:

H1: There will be a main effect of Group Membership. Those who anticipate costs from an in-group member will score lower on the Raven's than those who anticipate costs from an out-group member.

CHAPTER 9

STUDY TWO: ANALYSES

Subjects and General Procedures

Subjects were recruited from undergraduate classes at the University of South Carolina and assigned randomly to experimental conditions. Again, all subjects were prescreened to be white, right-handed females. Seventeen subjects participated in each condition and each subject was paid $10 for participation in the study. Procedures for this follow-up study mirrored that of the initial study, but differed only in regards to the manipulation detailed below.

Data

Data was excluded for one participant because of a lack of belief about the veracity of the manipulation. This left 34 analyzable cases. Of these cases all were female. All reported their race as white in the pre-study, however during the actual study, one person categorized themselves as White-Hispanic. We kept this cases in the analysis because of their previous reports and because this should not theoretically effect results.

Descriptive Statistics

Group membership represented the experimentally manipulated variable indicating whether participants thought that their in-group or out-group sanctioned high test performance. We also administrated a series of questions that probed into the level of

identification that participants had with their in-group. Principal components analysis indicated that these questions indicated a latent concept as 84.16% of the variance was modeled by the extracted component. We standardized the responses and combined them into a scale measure of level of group identification (see Table 9.1).

Table 9.1. Principal Components Analysis.

Group Identification Question	*Factor Loading*
See Myself as Part of the Group	.873
Pleased to be a Part of the Group	.915
Strong Ties to the Group	.939
Identify with the Group	.937
Group/Self Interconnectedness	.922

Given that intelligence tests are correlated with other academic measures, we collected information about subjects' educational background during pre-study questionnaires. They reported an estimate of their high school and college grade point averages and either their ACT or SAT scores. College admissions test scores were standardized to account for their different scales of measurement. Participants also recorded self-reports of both their mother and father's level of education. These reports were collapsed into binary categories indicating whether or not a college degree was obtained. Finally, we computed the time that each subject spent working on the test.

We analyzed the descriptives of each parameter. Table 9.2 presents the mean, standard deviation, skew and kurtosis of each variable. The Kolmogorov-Smirnov test showed that the dependent variable is normally distributed, D= .141, p=.084.

Table 9.2. Descriptive Statistics.

Parameter	*Mean (SD)*	*Skew*	*Kurtosis*
Independent Variables			
Group Membership	--- (---)	---	---
College Admissions Test Score	.000 (.953)	-.652	1.28
Effort on Test (TS_Total)	32.128 (7.71)	-.091	-.963
Level of Group Identification	.000 (4.478)	-1.53	1.707
College GPA	.863 (.103)	-.825	-.064
High School GPA	.888 (.094)	-1.107	1.263
Mother's Level of Education	.765 (---)	---	---
Father's Level of Education	.824 (---)	---	---
Dependent Variables			
Raven's Test Score	23.794 (4.637)	-.070	-.924

Regression Analysis

Regression Diagnostics

Before model building, we tested the data for collinearity, heteroscedasticity, and outliers. Variance inflation factors indicated no collinearity (see Table 9.3). The Breusch-Pagen test showed that the data were homoscedastic, x2=.002, p= 0.964. The Kolmogorov-Smirnov test showed that the dependent variable is normally distributed, D= .141, p=.084. Prior to model building, we ran an analysis for multivariate outliers on the full model using Cook's Distance. We eliminated two cases because they had a distance greater than 4/n, indicating their undue influence on the regression coefficients (Bollen and Jackman 1990).

Table 9.3. Collinearity Diagnostics.

Parameter	*Variance Inflation Factor*
Group Membership	1.29
Standardized Test Score	1.08
Effort on Test (TS_Total)	1.28

Note: Variance Inflation Factors less than 10 indicate no collinearity.

Model Building. A series of F-tests for nested models showed that constraining College and High School GPA, Mother's and Father's level of education, and identification with the group resulted in improved model fit, $F(4,25) = 0.311$, $p = 0.868$ and $F(1,29) = 1.817$, $p = 0.188$ (see Appendix I for calculations). We chose use this more parsimonious and better fitting model, despite that the results were robust to the inclusion of these variables because keeping these variables in the model would be equivalent to parameterizing noise (see Table 9.4).

Rather than including effort on the test as a mediator, we chose to include it as a control in this model. We did this because in the prior analysis, effort only mediated the relationship between perceived status and test score. In this study, there were no status differences between participants and, as such, we should not expect to find any differences in effort.

Final Model

Controlling for college admission test scores and effort on test, we found that those who anticipated that a high score on the test would result in both financial and social costs from a member of their in-group scored lower than if those costs were perceived to come from someone in their out-group. Table 9.5 shows the predicted marginal means per condition. Those who perceived in-group costs for success scored almost 2 points lower than those who thought that their costs came from an out-group member.

Table 9.4. F-Test of Nested Models: Study Two.

	Model 1 B (SE)	*Model 2* B (SE)	*Model 3* B (SE)
Intercept	14.557 (9.769)	10.142 (4.657)*	9.594 (4.702)*
Independent Variables			
Group Membership	4.089 (1.697)*	4.421 (1.504)**	4.431 (1.524)**
College Admissions Test Score	2.359 (.912)*	2.190 (.755)**	1.943 (.743)*
Effort on Test (TS_Total)	.194 (.107)+	.219 (.099)*	.235 (.100)*
Level of Group Identification	.217 (.184)	.209 (.155)	
College GPA	6.539 (8.371)		
High School GPA	-9.790 (9.867)		
Mother's Level of Education	-.148 (2.519)		
Father's Level of Education	.049 (2.548)		
Omnibus F	2.258+	4.647**	5.443**

*Note: *=p<.05, **=p<.01, ***=p<.001.*

Table 9.5. Predicted Marginal Means.

Manipulation	*Raven Test Score*
Perceived In-Group Costs	22.82 (4.69)
Perceived Out-Group Costs	24.76 (4.51)

CHAPTER 10

DISCUSSION AND CONCLUSIONS

We found that the academic performance of low status individuals is most prominently affected by situations where they perceive high achievement as costly. This is especially pronounced when achievement is sanctioned by in-group members. Findings from Study One indicated about a half of a standard deviation difference in tests scores between the highest and lowest scoring conditions. To illustrate what these findings mean relative to standard IQ norms, Figure 10.1 details both Raven's scores and their normed equivalents. Those who were made to feel low status and experienced costs for success scored, on average, 22.65. Whereas, everyone else scored, on average, 24.05. That is equivalent to losing about 6 points on the IQ. To emphasize how large these score differences are, Table 10.1 details other processes that similarly decrease scores alongside their respective score differences. Our findings indicate that perceiving yourself to be of relative low status and expecting costs associated with success is almost as detrimental to scores as taking the test without sleeping the night before, being chronically drunk, or becoming elderly (Linde and Bergstrom 1992; Ponton et al. 1996; Jones and Parsons 1971).

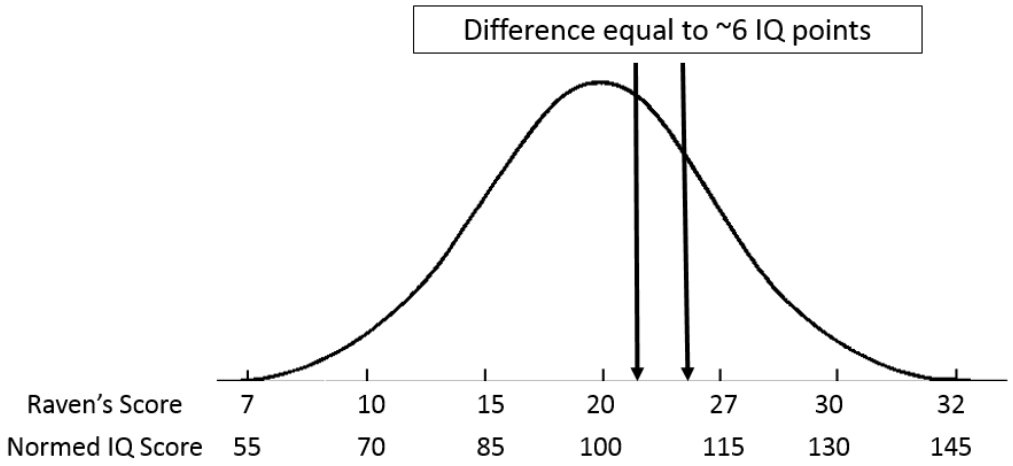


Figure 10.1 Raven's Scores in Relation to Normed IQs.

Table 10.1. Comparable Score Decrements and Their Processes.

Process	Approximate Score Decrement
Sleep deprivation	1 SD
Chronic alcohol consumption	1 SD
Becoming elderly	1 SD
Perceived low status and costs for achievement	½ SD

Effort as a Mediator

Reduced effort was the path by which low status subjects in our study underachieved. Unlike previous studies that identified test anxiety (a largely non-conscious phenomena) to lower scores, our study showed that students actively put forth less effort, if conditions were set such that there was a structural ceiling for their accomplishments. They both took less time to complete the test than high status others who had no such ceiling. This shows that the scores of low status individuals may not only not reflect their actual abilities, but that their underperformance could also be a reaction to larger social processes at play.

While behavioral indicators of effort show that effort mediated the relationship between status and test score, self-report measures of effort did not serve as a mediator in Study One. This is consistent with other work on socially motivated underachievement that would able to isolate relationships between behavioral indicators, but not self-report measures (White et al. 2002). This may have happened because while our measure of time spent taking the test occurred during the test, participant's filled out our self-report questionnaires after the experiment was over. Their memory of how hard they tried could have been compromised. In addition, there could have been a demand effect where they fictitiously inflated their level of effort to be in line with researcher expectations.

Crystallized and Fluid Intelligence

Our measure of intelligence, the Raven's, is one of only a few intelligence tests to be lauded as culturally non-dependent (Jensen 1980). This makes our findings even more pronounced. We were able to observe score decrements even under conditions where environmental factors should be the least effectual. Environmental factors, such as the social processes described throughout, are more important when the test is known to be culturally biased or when the subsets of the test measure crystallized intelligence (Kan et al. 2013). In classrooms and other testing environments, administrators typically use the WISC or WAIS tests to place students into special needs classrooms or to grant them admission to gifted and talented programs. These tests are comprised of subscales that measure both fluid and crystallized intelligence. As it stands, approximately half of the overall score on these tests are determined by measures that we know to be highly culturally dependent. Furthermore, the SAT and ACT, which greatly influence college

admissions decisions are purely measures of crystallized intelligence, as their scores are based on achievement rather than aptitude. They are, thus, the most culturally biased of all. Yet, despite that we know that the scores of disadvantaged test-takers are handicapped by these measures, we continue to administer these tests and use them to determine important life outcomes. This, in effect, stacks the cards squarely against those of low status. If we were able to induce score decrements of almost half of a standard deviation on a test that measures only fluid intelligence, it would follow that scores on these more culturally dependent tests would decrease much further.

Relevance for Status Interventions

Understanding how perceptions of social rewards and costs affect intelligence test performances is important for efforts to intervene in the status component. For instance, scholars have developed intervention strategies to lessen the association of low status with low ability. These promote equal influence in collective task settings by cancelling out the effects of status characteristics and have been shown to lessen or eliminate the effects of diffuse status characteristics such as gender (Lockheed and Hall 1976; Pugh and Wahrman 1983), race (Cohen and Roper 1972; Lohman 1972; Webster and Driskell 1978) and educational attainment (Markovsky et al. 1984), as well as specific characteristics such as reading ability and academic reputation (Tammivaara 1982; Rosenholtz 1977).

Ridgeway (1982) showed that when low status actors assert themselves as group-motivated rather than self-motivated, they were able to overcome the low-status handicap and attain high influence. For Pugh and Wahrman (1983), interventions either consisted

of verbal statements of equality, demonstrations of equality or demonstrations of superiority of low status actors. They found that only the low status actors who demonstrated their superior competence were able to lessen status effects. Additionally, Cohen and Roper (1972) found that status effects were only lessened when low status actors demonstrate their superior competence by serving as teachers of high status actors.

These interventions introduced new status information that effectively overcame low status, but still left unexplained the sources for the actual performance decrements that have been shown to be associated with low status. Essentially the interventions took broad, scattershot approaches to resolving the problem. Our research stands to provide a more refined understanding of the process, in turn allowing milder and more subtle interventions to be tested and applied in future research.

Future Research

This research improved our understanding of an underlying mechanism driving variability in mental ability test scores. We hope to continue this line of research in the future, building on the findings from this project in designing more efficient and subtle kinds of status interventions for increasingly naturalistic settings. Returning to one of our original motivations, this research addressed an anomalous finding regarding aptitude test performance for low status versus high status group members— the observation that, even in the absence of any real aptitude differences, some minority group members (African Americans) score lower than majority group members (Caucasians), while females score as well as males. Such findings may have been due to perceived rewards

and costs that were factors for the test-takers, but left unmeasured in the research settings that produced those findings.

The perception of higher costs associated with high performance in low status actors can be further investigated as a mediating factor between status and mental ability test scores. In fact, a complement to this study is in the works that shows how those of different statuses differentially perceive rewards and costs for outstanding academic achievement.

In addition to our follow-up studies, this research can be replicated in the lab and also confirmed in more natural settings. In the future, the assertion that low status actors perform under their ability on tests because of costs associated with high performance can be expanded to encapsulate other aspects of education. For example, does the perception of high costs have effects on other types of tests, such as the SAT and GRE? Do these results generalize to less critical test situations, such as a high-school biology midterm?

Our research design and procedures could be used in many future projects. For example, our study uses the Raven test because of its cultural neutrality. Our results would establish a baseline by which to compare the effects of perceived rewards and costs on more culturally biased tests. This would allow researchers to examine the extent to which cultural biased tests either exacerbate or ease status-related beliefs. Additionally, our design could apply more directly to research on stereotype threat. Future studies could manipulate the presentation of the mental ability test as either a test or a filler task. If there are differences between test- takers that are knowingly taking a mental ability test as opposed to thinking that they are taking a filler task, then

researchers could tell much of the effect is attributable to group-stereotypes verses status-beliefs. In short, this study could be the beginning of a program of research about the effects of status, costs and academic performance.

The more thorough our understanding of the mechanisms behind these phenomena, the better our ability to eliminate the arbitrary inequities and long-term disadvantages that can result from status processes that have little or nothing to do with the task at hand. This could be a key to lessening the achievement gap between high and low status actors. For this reason, it is important to isolate the underpinnings of status-based differences in test performance and then to develop methods for offsetting them.

REFERENCES

Ainsworth-Darnell, J., and Downey, D. "Assessing the Oppositional Culture Explanation for Racial/Ethnic Differences in School Performance." American Sociological Review 63(4):536-553.

Aronson, J., and Salinas, M. 1997. "Stereotype Threat, Attributional Ambiguity, and Latino Underperformance." Unpublished Manuscript: University of Texas.

Aronson, J., Fried, C., and Good, C. 2002. "Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence." Journal of Experimental Social Psychology (38):113-125.

Arroyo, C., and Zigler, E. 1995. "Racial Identity, Academic Achievement, and the Psychological Well- Being of Economically Disadvantaged Adolescents." Journal of Personality and Social Psychology 69:903–14.

Austen-Smith, D., and Fryer, R. 2005. "An Economic Analysis of 'Acting White'." The Quarterly Journal of Economics 120(2): 551-583.

Bales, R. 1950. Interaction Process Analysis: A Method for the Study of Small Groups. Oxford, England: Addison-Wesley.

Balkwell, J. 1991. "From Expectations to Behavior: An Improved Postulate for Expectations." American Sociological Review. 56(3):355-369.

Barbezat, D. 1987. "Salary Differentials by Sex in the Academic Labor Market." Journal of Human Resources 22:422-28.

Barr, R., and Dreeben, R. 1983. How Schools Work. Chicago: University of Chicago Press.

Belmont, L., and Marolla, F. 1973. “Birth Order, Family Size, and Intelligence: A Study of a Total Population of 19-year-old Men Born in the Netherlands is Presented.” Science 182(4117):1096-1101.

Benyamin, B., St Pourcain, B., Davis, O., Davies, G, Hansell, N., Brion, M., Kirkpatrick, R., Cents, R., Franić, S., Miller, M., Haworth, C., Meaburn, E., Price, T., Evans, D., Timpson, N., Kemp, J., Ring, S., McArdle, W., Medland, S., Yang, J., Harris, S., Liewald, D., Scheet, P., Xiao, X., Hudziak, J., de Geus, E., Welcome Trust Case Control Consortium, Jaddoe, V., Starr, J., Verhulst, F., Pennell, C., Tiemeier, H., Iacono, W., Palmer, L., Montgomery, G., Martin, N., Boomsma, D., Posthuma, D., McGue, M., Wright, M., Davey Smith, G., Deary, I., Plomin, R., and Visscher, P. “Childhood Intelligence is Heritable, Highly Polygenic and Associated with FNBP1L.” Molecular Psychiatry 19(2):253–258.

Berger, J. 1958. “Relations between Performance, Rewards and Action-Opportunities in Small Groups." Ph.D. Dissertation, Harvard University, Cambridge, MA.

Berger, J., Cohen, B., and Zelditch, M. 1966. "Status Characteristics and Expectation States." Pp. 29-46 in Sociological Theories in Progress, Vol. 1, edited by J. Berger, M. Zelditch Jr, and B. Anderson. Boston: Houghton Mifflin.

Berger, J., Cohen, B., and Zelditch, M. 1972. “Status Characteristics and Social Interaction.” American Sociological Review 37(3):241-255.

Berger, J., and Conner, T. 1974. “Performance Expectations and Behavior in Small Groups: A Revised Formulation.” Pp. 85-110 in Expectation States Theory: A Theoretical Research Program, edited by Joseph Berger, Thomas Conner, and M. Hamit Fisek. Winthrop.

Berger, J., Fisek, H., Norman, R., and Zelditch, M. 1977. Status Characteristics and Social Interaction. New York: Elsevier.

Berger, J., Fisek, H., Norman, R., and Wagner, D. 1998. “Formation of Reward Expectations in Status Situations.” Pp. 215-61 in Joseph Berger and Morris Zelditch, Jr., (eds.), Status, Rewards and Influence. San Francisco: Jossey-Bass.

Bergin, D., and Cooks, H. 2002. "High School Students of Color Talk About Accusations of ‘Acting White’.” The Urban Review 34:113-34.

Bertran, M. and Hallock, K. 2001 “The Gender Gap in Top Corporate Jobs” Human Resource Studies University of Chicago.

Biddle, J. Hamermesh, D. 1998. “Beauty, Productivity and Discrimination: Lawyers’ Looks and Lucre” Journal of Labor Economics 16:172-201.

Blau, P. and Duncan, O. 1967. The American Occupational Structure. John Wiley & Sons, Inc.: New York, N.Y.

Blau, F., and Kahn, L. 1994. “Rising Wage Inequality and the U.S. Gender Gap” The American Economic Review 84(2).

Bourdieu, Pierre. 1977. Outline of a Theory of Practice. Cambridge, England: Cambridge University Press.

Bouchard, T., Lykken, D., McGue, M., Segal, N., and Tellegen, A. 1990. "Sources of Human Psychological Differences: The Minnesota Study of Twins Reared Apart." Science 250:223–28.

Bowles, S., and Gintis, G. 2002. "Schooling in Capitalist America Revisited." Sociology of Education 75(1):1-18.

Bradley, R., and Corwyn, R. 2002. "Socioeconomic Status and Child Development." Annual Review of Psychology 53:371–99.

Brophy, J. 1983. "Research on the Self-Fulfilling Prophecy and Teacher Expectations." Journal of Educational Psychology 75(5):631-61.

Cadinu, M., Maass, A., Rosabianca, A., and Kiesner, J. 2005. "Why Do Women Underperform Under Stereotype Threat?: Evidence for the Role of Negative Thinking." Psychological Science 16:572.

Capron, C., and Duyme, M. 1989. "Assessment of Effects of Socio-economic Status on IQ in a Full Cross-fostering Study." Nature 340:552–54.

Carroll, J. 1993. Human Cognitive Abilities: A Survey of Factor Analytic Abilities. New York: Cambridge University Press.

Cattell, R. 1941. "Some Theoretical Issues in Adult Intelligence Testing. Psychological Bulletin 38:592.

Chabris, C., Hebert, B., Benjamin, D., Beauchamp, J., Cesarini, D., van der Loos, M., Johannesson, M., Magnusson, P., Lichtenstein, P., Atwood, C., Freese, J., Hauser, T., Hauser, R., Christakis, N., and Laibson, D. 2012. "Most Reported Genetic Associations with General Intelligence Are Probably False Positives." Psychological Science 23(11):1314-1323.

Cohen, E., and Roper, S. 1972. “Modification of Interracial Interaction Disability: An Application of Status Characteristics Theory.” American Sociological Review 37:648-55.

Cohen, G., Steele, C., and Ross, L. 1999. The Mentor’s Dilemma: Providing Critical Feedback across the Racial Divide.” Personality and Social Psychology Bulletin 25:1302-1318.

Collins, R. 1979. The Credential Society: An Historical Sociology of Education and Stratification. John Wiley, New York.

Cook, P., and Ludwig, J. 1997. “Weighing the “Burden of ‘Acting White’”: Are There Race Differences in Attitudes toward Education?” Journal of Policy Analysis and Management 16(2):256–278.

Croizet, J., and Claire, T. 1998. Extending the Concept of Stereotype Threat to Social Class: The Intellectual Underperformance of Students from Low Socioeconomic Backgrounds. Personality and Social Psychology Bulletin 24:588–594.

Daley, T., Whaley, S., Sigman, M., Espinosa, M., and Neumann, C. 2003 “IQ on the Rise: The Flynn Effect in Rural Kenyan Children.” Psychological Science 14(3):215-219.

Datnow, A., and Cooper, R. 1997. Peer Networks of African American Students in Independent Schools: Affirming Academic Success and Racial Identity. Journal of Negro Education 65(4):56-72.

David, B., and Turner, J. 1996. “Studies in Self-categorization and Minority conversion: Is Being a Member of the Out-group an Advantage.” British Journal of Social Psychology 35, 179-199.

David, B., and Turner, J. 2001. "Self-categorization Principles Underlying Majority and Minority Influence." In J. P. Forgas & K. D. Williams (Eds.), Social influence: Direct and indirect processes. Philadelphia, PA: Psychology Press.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S., Liewald, D., Ke, X., Le Hellard, S., Christoforou, A., Luciano, M., McGhee, K., Lopez, L., Gow, A., Corley, J., Redmond, P., Fox, H., Haggarty, P., Whalley, L., McNeill, G., Goddard, M., Espeseth, T., Lundervold, A., Reinvang, I., Pickles, A., Steen, V., Ollier, W., Porteous, D., Horan, M., Starr, J., Pendleton, N., Visscher, P., and Deary, I. 2011. "Genome-wide Association Studies Establish that Human Intelligence is Highly Heritable and Polygenic." Molecular Psychiatry 16:996-1005.

Deary, I. 2008. "Why Do Intelligent People Live Longer?" Nature 456:175-176.

Deary, I, Whalley, L., Lemmon, H., Crawford, J., and Starr, J. 2000. The Stability of Individual Differences in Mental Ability from Childhood to Old Age: Follow-up of the 1932 Scottish Mental Survey. Intelligence. 28(1):49-55.

Deary, I., Johnson, W., and Houlihan, L. "Genetic Foundations of Human Intelligence." Human Genetics 126:215–232.

Desert, M., Preaux, M., and Jund, R. 2009. So Young and Already Victims of Stereotype Threat: Socio-economic Status and Performance of 6 to 9 Years Old Children on Raven's Progressive Matrices. European Journal of Psychology of Education 24:207-218.

Dippong, J. 2012. "The Effects of Scope Condition-Based Participant Exclusion on Experimental Outcomes in Expectation States Research: A Meta-Analysis." Social Science Research 41(2):359-371.

Duyme, M., Dumaret, A., and Tomkiewicz, S. 1999. "How Can We Boost IQs of 'Dull Children'?: A Late Adoption Study." Proceedings of the National Academy of Science 96(15):8790–8794.

Eccleston, M. 2011. In Utero Exposure to Maternal Stress: Effects of the September 11th Terrorist Attacks in New York City on Birth and Early Schooling Outcomes. Cambridge, MA: Harvard University.

Exline, J., Zell, A., and Lobel, M. 2013. "Sidestepping Awkward Encounters: Avoidance As a Response to Outperformance-related Discomfort." Journal of Applied Social Psychology 43:706–720.

Exline, J., and Lobel, M. 1999. The Perils of Outperformance: Sensitivity about Being the Target of a Threatening Upward Comparison. Psychological Bulletin, 125, 307-337.

Exline, J., and Lobel, M. 2001. Private Gain, Social Strain: Do Relationship Factors Shape Responses to Outperformance? European Journal of Social Psychology 31:593-607.

Exline, J., Single, P, Lobel, M., and Geyer, A. 2004. "Glowing Praise and the Envious Gaze: Social Dilemmas Surrounding the Public Recognition of Achievement." Basic and Applied Social Psychology 26:119-130.

Farkas, G., and Vicknair, K. 1996. "Appropriate Test of Racial Wage Discrimination Require Controls for Cognitive Skill: Comment on Cancio, Evans, and Maume." American Sociological Review 61(4):557-60.

Feinstein, L. 2003. "Inequality in the Early Cognitive Development of British Children in the 1970 Cohort." Economica 70(277):73-97.

Ferguson, R. 2001. "A Diagnostic Analysis of Black-White GPA Disparities in Shaker Heights, Ohio." Pp. 347-414 in Brookings Papers on Education Policy 2001, edited by Diane Ravitch. Washington, DC: Brookings Institution Press.

Fischer, C., Hout, M., Jankowski, S., Lucas, M., S., Swidler, A, Voss, K. 1996. Inequality by Design: Cracking the Bell Curve Myth. Princeton University Press, Princeton, NJ.

Flynn, J. 1987. "Massive IQ Gains in 14 Nations: What IQ Tests Really Measure." Psychological Bulletin 101:171–191.

Flynn, J. 2007. What is Intelligence? Beyond the Flynn Effect. New York, NY: Cambridge University Press.

Fordham, S., and Ogbu, J. 1986. "Black Students' School Success: Coping with the Burden of 'Acting White.'" Urban Review 18:176–206.

Forste R., Weiss J., and Lippincott E. 2001. The Decision to Breastfeed in the United States: Does Race Matter?" Pediatrics 108:291–296.

Foschi, M. 1996. "Gender and Double Standards in the Assessment of Job Applicants." Social Psychology Quarterly 57:326-39.

Fryer, R., and Torelli, P. 2010. "An Empirical Analysis of 'Acting White'." Journal of Public Economics 94:380-396.

Garnoran, A., and Mare, R. 1989. Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality? American Journal of Sociology 94:1146-1183.

Gander, J. 1997. "Gender-Based Faculty-Pay Differences in Academe: A Reduced-Form Approach." Journal of Labor Research 18(3):451–61.

Geary, D., Saults, S. Liu, F., and Hoard, M. 2000. “Sex Differences in Spatial Cognition, Computational Fluency, and Arithmetical Reasoning.” Journal of Experimental Child Psychology 77:337-353.

Gerhart, B. and Milkovich, G. 1989. “Organizational Differences in Managerial Compensation and Financial Performance.” Working Paper Center for Advanced Human Resource Studies, Cornell University.

Good, C., Aronson, J., and Harder, J. 2008. “Problems in the Pipeline: Stereotype Threat and Women's Achievement in High-level Math Courses.” Journal of Applied Developmental Psychology 29:17–28.

Gottfredson, L. 2004 “Intelligence: Is It the Epidemiologists’ Elusive ‘Fundamental Cause’ of Social Class Inequalities in Health?” Journal of Personality and Social Psychology 86(1):174–199

Gottfredson, L. 1997. Mainstream Science on Intelligence: An Editorial with 52 Signatories, History, and Bibliography. Intelligence (24):13-23.

Gow, A., Johnson, W., Pattie, A., Brett, C. E., Roberts, B., Starr, J. M., and Deary, I. J. 2011. “Stability and Change in Intelligence from Age 11 to Ages 70, 79, and 87: the Lothian Birth Cohorts of 1921 and 1936.” Psychological Aging 26:232-240.

Griliches, Z., and Mason, W. 1972. “Education, Income and Ability.” Journal of Political Economy 80:S74–S103.

Grodsky, E., and Pager, D. 2001. “The Black–White Wage Gap: Occupational and Individual Sources.” American Sociological Review 66:542–567.

Hanushek, E., and Woessmann, L. 2006. “Does Early Tracking Affect Educational Inequality and Performance? Differences in Differences Evidence across Countries.” Economic Journal 116(510):C63–C76.

Hall, M., and Farkas, G. 2011. “Adolescent Cognitive Skills, Attitudinal/Behavioral Traits and Career Wages.” Social Forces 89(4):1261-1285.

Harrod, W. J. 1980. “Expectations from unequal rewards.” Social Psychology Quarterly 43:126-30.

Hart, B., and Risley, T. 1995. Meaningful Differences in the Everyday Experience of Young American Children. Baltimore: Paul H. Brookes.

Hartigan, J., and Wigdor, A. 1989. Fairness in employment testing. Washington, DC: National Academy Press.

Hauser, R., Tsai, S., and Sewell, W. 1983. “A Model of Stratification with Response Error in Social and Psychological Variables.” Sociology of Education 56:20-46.

Hembree, R. 1988. “Correlates, Causes, Effects, and Treatment of Test Anxiety.” Review of Educational Research 58(1):47-77

Herrnstein, R. and Murray, C. 1994. The Bell Curve. New York: Free Press.

Homans, G. 1961. Social Behavior: Its Elementary Forms. Oxford, England: Harcourt, Brace.

Horn, J. 1965. Fluid and Crystallized Intelligence. Urbana Champaign: University of Illinois.

Hunter, J. 1983. Validity Generalization of the ASVAB: Higher Validity for Factor Analytic Composites. Rockville, MD: Research Applications.

Hunter, J. 1986. Cognitive Ability, Cognitive Aptitudes, Job Knowledge, and Job Performance. Journal of Vocational Behavior 29:340-362.

Hunter, J., and Schmidt, F. 1996. Intelligence and Job Performance: Economic and Social Implications. Psychology, Public Policy, and Law 2:447-472.

Jemmott, J., and E. Gonzalez. 1989. "Social Status, the Status Distribution, and Performance in Small Groups." Journal of Applied Social Psychology 19:584–598.

Jencks, C., Bartlett, S., Corcoran, M., Crouse, J., Eaglesfield, D., Jackson, G.1979. Who Gets Ahead? The Determinants of Economic Success in America. New York: Basic Books, Inc.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H. 1972. Inequality: A Reassessment of the Effect of Family and Schooling in America. New York: Basic Books, Inc.

Jennings, J., Deming, D., Jencks, C., Lopuch, M., and Schueler, B. 2015. "Do Differences in School Quality Matter More Than We Thought? New Evidence on Educational Opportunity in the Twenty-first Century." Sociology of Education 88(1):56–82.

Jensen, A. 1969. "How Much Can We Boost IQ and Scholastic Achievement?" Harvard Educational Review 39:1-123.

Jensen, A. 1974. "How Biased are Culture-loaded Tests?" Genetic Psychology Monographs 90:185-244.

Jensen, A. 1980. Bias in Mental Testing. New York: Free Press.

Jensen, A. 1992. Spearman's Hypothesis: Methodology and Evidence. Multivariate Behavioral Research 27:225-234.

Jensen, A. 1998. The g Factor: The Science of Mental Ability. Westport, CT: Praeger Publishers.

Jones, B., and Parsons, O. 1971. "Impaired Abstracting Ability in Chronic Alcoholics." Archives of General Psychiatry 24(1):71-75.

Kahneman, D. and Tversky, A. 1979. Intuitive Prediction: Biases and Corrective Procedures. TIMS Studies in Management Science 12:313-327.

Kalkhoff, W., and S. Thye. 2006. "Expectation States Theory and Research: New Observations from Meta-analysis." Sociological Methods and Research 35(2):219-49.

Kan, K., Wicherts, J., Dolan, C., and van der Maas, H. 2013. "On the Nature and Nurture of Intelligence and Specific Cognitive Abilities: The More Heritable, the More Culture Dependent." Psychological Science 24(12):2420-2428.

Katz, L., Murphy, K. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." The Quarterly Journal of Economics 107(1):35-78.

Kerckhoff A. 2001. "Education and Social Stratification Processes in Comparative Perspective." Sociology of Education 74:3–18.

Lanphear, B., Hornung, R., Ho, M., Howard, C., Eberly, S., Knauf, K. 2002. "Environmental Lead Exposure during Early Childhood." The Journal of Pediatrics 140(1):40-47.

Lareau, A. 2003. Unequal Childhoods: Race, Class, and Family Life. Berkeley: University of California Press.

Linde, L., and Bergstrom, M. 1992. "The Effect of One Night without Sleep on Problem-solving and Immediate Recall." Psychological Research 54:127-136.

Lockheed, M., and Hall, K. 1976. "Conceptualizing Sex as a Status Characteristic: Applications to Leadership Training Strategies." Journal of Social Issues 32(3): 111-124.

Lohman, M. 1972 "Changing a Racial Status Ordering Implications for Desegregation." Journal of Educational and Urban Society 4:383-402.

Lovaglia, M., and Lucas, J. 1997. "Group Processes and Individual Scores on Standardized Tests: A Theoretical Note and Basis for Investigation" Current Research in Social Psychology 2:1–13.

Lovaglia, M., Lucas, J., Houser, J., Thye, S., and Markovsky, B. 1998. "Status processes and Mental Ability Test Scores" American Journal of Sociology 104 (1):195–228.

Lovaglia, M., Youngreen, R., Lucas, J., Nath, L., Rutstrom, E., and Willer, D. 2004. "Stereotype Threat and Differential Expected Consequences: Explaining Group Differences in Mental Ability Test Scores." Sociological Focus 37:2:107-125.

Lucas, J. 1999. "Behavioral and Emotional Outcomes of Leadership in Task Groups." Social Forces 78:747-776.

Lucas S. 2001. "Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects." American Journal Sociology 106:1642-90.

Lynn, R. 1990. "The Role of Nutrition in Secular Increases in Intelligence." Personality and Individual Differences 11:273-285.

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. 2013. “Poverty Impedes Cognitive Function.” Science. 341:976.

Markovsky, B., L. Smith and J. Berger. 1984. “Do Status Interventions Persist?” American Sociological Review 49:373-382.

McGlone, M., and Aronson, J. 2006. “Stereotype Threat, Identity Salience, and Spatial Reasoning.” Journal of Applied Developmental Psychology 27:486-493.

McGrew, K. and Knopik, S. 1993. “The Relationship between the WJ–R Gf-Gc Cognitive Clusters and Writing Achievement across the Life-span.” School Psychology Review 22:687-695.

Mingroni, M. 2004. “The Secular Rise in IQ: Giving Heterosis a Closer Look.” Intelligence 32:65-83.

Mitra A., Khoury A., Hinton A., Carothers C. 2004. “Predictors of Breastfeeding Intention among Low-income Women.” Maternal and Child Health Journal 8:65-70.

Moore, E. 1986. “Family Socialization and the IQ Test Performance of Traditionally and Transracially Adopted Black Children.” Developmental Psychology 22:317-326.

Morgan, L. A. 1998. Glass-ceiling Effect or Cohort Effect? A Longitudinal Study of the Gender Earnings Gap for Engineers, 1982 to 1989. American Sociological Review 63(4): 479-83.

Morgan L. 1998. “Glass-ceiling Effect or Cohort Effect? A Longitudinal Study of the Gender Earnings Gap for Engineers, 1982 to 1989.” American Sociological Review 63:479-93.

Mortensen, E., Michaelsen, K., Sanders, S., and Reinisch, J. 2002. "The Association between Duration of Breastfeeding and Adult Intelligence." Journal of the American Medical Association 287:2365-71.

Neal, D., and Johnson, W. 1996. "The Role of Premarket Factors in Black-White Wage Differences." Journal of Political Economy 104(5):869-895.

Needleman, H. L., Geiger, S.K., and Frank, R. 1985. "Lead and IQ Scores: A Reanalysis." Science 227:701–4.

Neisser, U., Boodoo, G., Bouchard, T., Boykin, A.W., Brody, N., Ceci, S., Halpern, D., Loehlin, J., Perloff, R., Sternberg, R., and Urbina, S. 1996. "Intelligence: Knowns and Unknowns." American Psychologist 51:77–101.

Nisbett, R. 2009. Intelligence and How to Get It: Why Schools and Cultures Count. New York: Norton.

Nisbett, R., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., and Turkheimer, E. 2012. "Intelligence: New Findings and Theoretical Developments." American Psychologist 67:130-159.

O'Neill, J. 1990. "The Role of Human Capital in Earnings Differences between Black and White Men." Journal of Economic Perspectives 4(4):25-45.

Oakes, J. 1992 "Can Tracking Research Inform Practice? Technical, Normative, and Political Considerations" Educational Researcher 21(4):12−21.

Olson, C. and Becker, B. 1983. "Sex Discrimination in the Promotion Process." Industrial and Labor Relations Review 36(4):624-641.

Osborne, J. 1995. "Academics, Self-esteem, and Race: A Look at the Assumptions Underlying the Disindentification Hypothesis." Personality and Social Psychology Bulletin 21:449-455.

Osborne, J. 2006. "Gender, Stereotype Threat and Anxiety: Psychophysiological and Cognitive Evidence." Journal of Research in Educational Psychology 8:109-138.

Osborne, J. 2007. "Linking Stereotype Threat and Anxiety." Educational Psychology 27:135-154.

Plomin, R., Hill, L., Craig, I., McGuffin, P., Purcell, S., Sham, P., Lubinski, D., Thompson, L., Fisher, P., Turic, D., and Owen, M. 2001 "A Genome-Wide Scan of 1842 DNA Markers for Allelic Associations with General Cognitive Ability: A Five-Stage Design Using DNA Pooling and Extreme Selected Groups." Behavior Genetics 31(6).

Ployhart, R., Ziegert, J., and McFarland, L. 2003. "Understanding Racial Differences on Cognitive Ability Tests in Selection Contexts: An Integration of Stereotype Threat and Applicant Reactions Research." Human Performance 16:231-259.

Pollitt, E., Gorman, K., Engle, P., Martorell, R., and Rivera, J. 1993. "Early Supplemental Feeding and Cognition." Monographs of the Society for Research in Child Development 58(7).

Pontón, M., Satz, P., Herrera, L., Ortiz, F., Urrutia, C., Young, R., D'Elia, L., Furst, C., and Namerow, N. 1996. "Normative Data Stratified by Age and Education for the Neuropsychological Screening Battery for Hispanics (NeSBHIS): Initial Report." Journal of International Neuropsychology Society 2:96-104.

Pugh, M., and Wahrman, R. 1983. “Neutralizing Sexism in Mixed-Sex Groups: Do Women Have to Be Better than Men?” American Journal of Sociology 88:746-62.

Putnam, R. 1995 “Bowling Alone: America’s Declining Social Capital” Journal of Democracy 6(1):65-78.

Raven, J. 1938. Progressive Matrices: A Perceptual Test of Intelligence. London: H.K. Lewis

Raven, J. C., Court, J., and Raven, J. 1992. “Standard Progressive Matrices.” Raven Manual: Section 3. Oxford: Oxford Psychologists Press.

Ridgeway, C. 1982. “Status in Groups: The Importance of Motivation.” American Sociological Review 47:76–88.

Rohde, T. and Thompson, L. 2007. “Predicting Academic Achievement with Cognitive Ability.” Intelligence 35:83-92.

Rosenholtz, S. 1977 “The Multiple Ability Curriculum: an Intervention against the Self-fulfilling Prophesy." Ph.D. Dissertation, School of Education, Stanford University.

Rosenthal, R., and Jacobson, L. (1968) 1992. Pygmalion in the Classroom. New York: Irvington.

Rowe, D., Jacobson, K., and Van den Oord, E. 1999. “Genetic and Environmental Influences on Vocabulary IQ: Parental Education Level as a Moderator.” Child Development 70:1151-1162.

Rubovits, P., and Maehr, M. 1973. “Pygmalion in Black and White.” Journal of Personality and Social Psychology 25:210–18.

Rushton, J. P., Skuy, M., and Bons, T. A. 2004. “Construct Validity of Raven's Advanced Progressive Matrices for African and Non-African Engineering Students in South Africa”. International Journal of Selection and Assessment 12:220−229.

Scarr, S. 1997. Behavior-genetic and Socialization Theories of Intelligence: Truce and Reconciliation. In R. J. Sternberg & E. Grigorenko (Eds.), Intelligence, heredity, and environment: 3-41. Cambridge: Cambridge University Press.

Scarr-Salapatek, S. 1971. “Race, Social Class, and IQ.” Science 174:1285-1295.

Schiff, M., Duyme, M., Dumaret, A., Stewart, J., Tomkiewicz, S., and Feingold, J. 1978. “Intellectual Status of Working-Class Children Adopted Early into Upper-Middle-Class Families.” Science 200:1503-4.

Schimel, J., Arndt, J., Banko, K., and Cook, A. 2004. “Not All Self-affirmations were Created Equal: The Cognitive and Social Benefits of Affirming the Intrinsic (vs. Extrinsic) Self.” Social Cognition 22:75-99.

Schmader, T. 2002. “Gender Identification Moderates Stereotype Threat Effects on Women’s Math Performance.” Journal of Experimental Social Psychology 38:194-201.

Schoon, I., Jones, E., Cheng, H., and Maughan, B. 2012. “Family Hardship, Family Instability, and Cognitive Development.” Journal of Epidemiology and Community Health 66:716-722.

Sewell, W., Haller, A., and Portes, A. 1969 “The Educational and Early Occupational Attainment Process.” American Sociological Review 34:82-92.

Simon, H. 1955. “A Behavioral Model of Rational Choice” The Quarterly Journal of Economics 69(1): 99-118.

Spearman, C. 1927. The Abilities of Man. New York: MacMillan.

Steele, C. 1992. “Race and the Schooling of Black Americans.” Atlantic Monthly 69.

Steele, C. 1997. “A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance.” American Psychologist 52:613-629.

Steele, C., and Aronson, J. 1995. “Stereotype Threat and the Intellectual Test Performance of African-Americans.” Journal of Personality and Social Psychology 69:797–811.

Steinberg, L., Dornbusch, S., and Brown, B. 1992. “Ethnic Differences in Adolescent Achievement.” American Psychologist 47:723–29.

Sternberg, R. 2012. “Intelligence.” Dialogues in Clinical Neuroscience 14(1):19-27.

Stone, J. 2002. “Battling Doubt by Avoiding Practice: The Effect of Stereotype Threat on Self-handicapping in White Athletes.” Personality and Social Psychology Bulletin 28:1667-1678.

Streissguth, A., Barr, H., and Sampson, P. 1990. “Moderate Prenatal Alcohol Exposure: Effects on Child IQ and Learning Problems at Age 71/2 Years.” Alcoholism: Clinical and Experimental Research 14:662–69.

Strenze T. 2007. “Intelligence and Socioeconomic Success: A Metaanalytic Review of Longitudinal Research.” Intelligence 35:401-426.

Stumpf, H., and Stanley, J. C. 1996. Gender-related Differences on the College Board’s Advanced Placement and Achievement Tests, 1982-1992. Journal of Educational Psychology 88: 353-364.

Sundre, D., and Moore, D. 2002. “The Student Opinion Scale: A Measure of Examinee Motivation.” Assessment Update 14(1):8-9.

Tajfel, H. 1979. “Individuals and Groups in Social Psychology.” British Journal of Social and Clinical Psychology 18:183-190.

Tammivaara, J. 1982 “The Effects of Task Structure on Beliefs about Competence and Participation in Small Groups.” Sociology of Education 55:212-22.

Terman, L., and Merrill, M. 1937. Measuring Intelligence. Houghton, Mifflin, New York.

Troyer, L. 2001. “Effects of Protocol Differences on the Study of Status and Social Influence.” Current Research in Social Psychology 6:182-205.

Turkheimer, E., Haley, A., Waldron, M., D’Onofrio, B., and Gottesman, I. 2003. “Socioeconomic Status Modifies Heritability of IQ in Young Children.” Psychological Science, 14(6):623-628.

Turner, J. 1985. “Social Categorization and the Self-concept: A Social Cognitive Theory of Group Behavior.” In E. J. Lawler (Ed.), Advances in Group Processes. Greenwich, CT: JAI Press.

Turner, J., Hogg, M., Oakes, P., Reicher, S., and Wetherell, M. 1987. Rediscovering the Social Group: A self-categorization Theory. Oxford, England: Basil Blackwell.

Tversky, A. and Kahneman, D. 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” Journal of Risk and Uncertainty 5: 297-323.

Tyree, H., Hicks, R. 1988. “Sex and the Second Moment of Prestige Distributions” Social Forces 66.

von Stumma, S., Plomin, R. 2015. “Socioeconomic Status and the Growth of Intelligence from Infancy through Adolescence.” Intelligence 48:30-36.

Wagner, D., Ford, R., and Ford, T. 1986. “Can Gender Inequalities Be Reduced?” American Sociological Review 51:47-61.

Walker, H., Ilardi, B., McMahon, A., and Fennell, M. 1996. “Gender, Interaction, and Leadership” Social Psychology Quarterly 59:255-72.

Webster, M., and Driskell, J. 1978. “Status Generalization: A Review and Some New Data” American Sociological Review 43:220-36.

Wechsler, D. 1939. The measurement of Adult Intelligence. Baltimore, MD: William and Wilkins.

White, K. 1982. “The Relation between Socioeconomic Status and Academic Achievement.” Psychological Bulletin 91:461–81.

White, P., Sanbonmatsu, D., Croyle, R., and Smittipatana, S. 2002. “Test of Socially Motivated Underachievement: ‘Letting Up’ for Others.” Journal of Experimental Social Psychology 38: 162-169.

Wise, S. and Kong, X. 2005. “Response Time Effort: A New Measure of Examinee Motivation in Computer-based Tests.” Applied Measurement in Education 18(2):163-183.

Wood, R, G,, Gorcoran, M, E,, & Gourant, P. N, 1993. “Pay Differences Among the Highly Paid: The Male-Female Earnings Gap in Lawyers' Salaries. Journal of Labor Economics, 11: 417-441.

Yang, J., Benyamin, B., McEvoy, B., Gordon, S., Henders, A., Nyholt, D., Madden, P., Heath, A., Martin, N., Montgomery, G., Goddard, M., Visscher P. “Common SNPs Explain a Large Proportion of the Heritability for Human Height. Nature Genetics 42:565-569.

Youngreen, R., Conlon, B., Robinson, D., and Lovaglia, M. 2009. “Identity Maintenance and Cognitive Test Performance.” Social Science Research 38:438-446.

APPENDIX A: PRE-TESTS

Pre-testing is the use of small pilot studies that allow the researcher to gauge, amongst other things, the precision of survey instruments, the flow of procedures, and the appropriateness of their experimental manipulations. In this case, we pre-tested a number of factors that could be influential to the outcome of the main studies. First, we wanted to ensure that our dependent variable, the Ravens Progressive Matrices, was appropriate for the population. We ran a pilot study of two versions of this test: the standard and the advanced. Using measurement theory, we were able to discriminate between test versions and choose the version most suitable to the main study's population. The function of the second pre-test was to assure us that one of the main experimental manipulations functioned as expected. We ran a pre-study that measured the effect of the handedness manipulation on the display of status-oriented behavior. We found that this techniques was, indeed, successful at producing the desired effects. Below, we detail the rationale behind the use of each pre-test, the procedures that we used, the outcomes, and their implications for the main studies.

Pre-test One: Ravens Progressive Matrices Test Version

The Raven Progressive Matrices (the "Ravens") provides a reliable and valid measure of mental ability and has been shown to correlate well with other tests of general intelligence (Raven, Court and Raven 1992; Jensen 1992). However, unlike other intelligence measures, it uses non-verbal questions. This suppresses any effects that may

have arisen from language deficits. Non-verbal items also make the Ravens test less culture-bound and thus less biased against minorities (Jensen 1974).

During the testing session, each subject is asked to identify a missing element that completes a pattern (see Appendix D for an illustration). The patterns get increasingly difficult as the subject moves through the test. They must complete the Ravens within a given time frame that varies with each version of the test. The two versions of interest to this study are the Ravens Standard Progressive Matrices (the "Standard") and the Ravens Advanced Progressive Matrices (the "Advanced"). The Standard is composed of 5 sets of 12 questions, totaling 60 potential points. This test is typically administered to children and adults of average intelligence. The Advanced, on the other hand, contains 36 total questions. Test administrators have marketed this version of the Ravens for adults who score near perfect scores on the Standard, making the Advanced the test of choice for those of above-average intelligence.

Scholars have used the Ravens to study sociological outcomes, such as educational potential and achievement, gender differences in spatial reasoning, and birth order and intelligence (Rohde and Thompson 2006; Geary, Saults, Liu and Hoard 2000; Belmont and Morolla 1973). However, to date, only two studies in the group processes research tradition have utilized the Ravens as a dependent variable. Lovaglia et al. (1998) used the Standard to determine how social status affected performance on intelligence tests. They noted, however, that the scores obtained using the Standard may be prone to a ceiling effect. In one of the three studies reported, subjects scored an average of 56 out of 60 points. While only two participants received perfect scores, it seems that such large range restriction might have detrimental effects on hypothesis testing. Aware of the

potential problems with the Standard, Youngreen et al. (2009) used the Advanced as the outcome measure in their study about identity maintenance and test performance. While they reported none of the problems occurring in the Lovaglia et al. (1998) study, their study did not provide much descriptive information about the test scores.

Since we had no conclusive reason to assume that one test was more suited for our main study than the other, we pre-tested both versions of the Ravens in our population. The test that we ultimately decided to use would need to satisfy the following requirements. First, it would need to produce a normal distribution. Most statistical tests require that cases sampled on the dependent variable come from a normal distribution. This means that the skewness and kurtosis statistics should be close to 0, cases examined with a Q-Q Plot should fall near the plotted line, and 99.9% of standardized scores should fall between -3.30 and 3.30. While there are, of course, statistical techniques to correct for non-normal distributions, it is always best to choose a dependent variable that needs the fewest transformations.

Secondly, the chosen test would need to produce the most variability. Tests that do not have the potential to produce variable scores on the dependent variable are poor indicators of the construct being measured. Tests that do produce high variability will have a large range of scores and a high variance coefficient. Variance statistics alone may be misleading, however. Since total variance is equal to the sum of both true score and error variance, a test may have a high variance simply because it contains a large amount of measurement error. To ensure this is not the case, the chosen test must also show a higher proportion of variance due to true score rather than error variance. This can be ascertained by examining the components of the KR-20 reliability statistic.

Finally, in addition to variability on the outcome, a good test will have maximum item variance, as well. This means that subjects will have about a 50/50 chance of answering any given question correctly. Higher percentages of per question correct scores may indicate a ceiling effect while lower percentages show the potential for a floor effect.

Design and Method

To test which version of the Ravens was more suitable for our main study, we employed a two condition between-subjects completely randomized design. Twenty Caucasian participants (12 females) were recruited from lower-level undergraduate classes at the University of South Carolina. Participants were randomly assigned to complete a computerized version of either the Standard or the Advanced. Upon completion of the test, subjects were debriefed about the nature of the study and paid $10 for their participation.

Results and Discussion

We computed test score for each condition by giving one point per correct answer. The mean score for the Standard was 48 (SD = 5.312) out of a potential 59 points. The distribution was negatively skewed (-.239, SE = .687) with a kurtosis of .331 (SE = 1.334). The mean score for the Advanced was 19.20 (SD = 7.223) out of a potential 36 points. The distribution had a positive skew (.217, SE = .687) and a kurtosis of -.627 (SE = 1.334). We standardized scores for each test version and discovered that no scores in either condition lay outside of the -3.3 to 3.3 range. Data points in each condition varied from the line in Q-Q plots, but with such a small sample size, it was difficult to draw

conclusions from this graph alone. The Shapiro-Wilk test of normality indicated that scores on both the Standard and the Advanced were drawn from a normal distribution (.958, p = .764, .957, p = .865, respectively). Since both tests produce a normal distribution, we will focus on their differences in variance.

To test which version of the Ravens produced the most variability, we first looked at the range of scores. Out of a possible 59 points, scores on the Standard ranged from 38 to 56 points. Scores ranged between 9 and 32 points out of a possible 36 points on the Advanced. This shows that the Advanced produces a broader range of scores in our population, despite its scale being structurally restricted because of the smaller number of items. This is further indicated by the variance of each test; the Standard had a lower variance ($\sigma 2 = 28.222$) than the Advanced ($\sigma 2 = 52.178$), despite having more items. Further analysis of the variance components computed using KR-20 indicates that 69.47% of the total variance of scores on the Standard was due to true score variance, whereas 85.19% of the score variance of the Advanced was attributed to true score differences. This not only means that the Advanced is a more reliable indicator, but that the variance on this test was not inflated because of measurement error.

To further discriminate between versions, we looked at the item variances of each test. Again, a good test will produce item variances of .5, indicating an item correct to item incorrect ratio of 50:50. On average, subjects answered 81% of Standard items correctly, whereas subjects answered 53% of Advanced items correctly. This shows that subjects were more likely to produce variable scores on each item when taking the Advanced rather than the Standard version of the Ravens.

Conclusions

Different versions of the Ravens have been used across disciplines to measure intelligence. Each version has its own caveats that can have consequences on research. For this reason, it is always best to know which version of the test will best suit the sample population. Our analysis shows that the Advanced version of the Ravens produces a normal distribution, a larger variance due to true score differences and larger item variances than the Standard version of the Ravens. This makes it a much better indicator of intelligence for our population.

Pre-test Two: Status Manipulation

The expectation states research program is known for its theoretical precision and methodological standardization. The wide-spread use of similar procedures in a standardized experimental setting ensures that findings relate both to one another and to the theory which spawned them. Within this tradition, status manipulations have been highly standardized. However, with novel status questions comes the use of novel status manipulations. How does the researcher make sure that the novel manipulation does what it is intended to do?

In this part of the appendix, we detail the background of Status Characteristics Theory as it relates to novel status manipulations. Then, we turn to an analysis of a new status manipulation: the handedness manipulation. We look to answer the following question: when utilizing a novel status manipulation, how do you know if it is successful in invoking status differences? Should you rely on the observed between condition relational differences or a comparison of observed and predicted values? We argue that

each approach yields useful information and then we provide evidence that reconciles the divide between these approaches.

Status Characteristics Theory

Inspired by the work of Bales (1950), Berger and colleagues (1966; 1972; 1977) developed a general theoretical account explaining how an individual's seemingly arbitrary personal characteristics can lead to them having more influence over a group. This theory, Status Characteristics Theory (SCT), asserts that when group members are oriented toward a collective and valued task, certain characteristics will differentiate the group members and ultimately affect the status hierarchy of the group.

The Standard Experimental Setting. In addition to the graph theoretic model that is used to make predictions about the relative status advantage of actors, SCT has developed a standardized experimental setting (SES) that is used to test these predictions. Typically, the setting consists of two individuals working collectively on a task. Status characteristics are manipulated by leading participants to believe that they possess more or less of an ostensibly genuine ability than their partner. This can be done by way of manipulating innate characteristics, such as "older" or "younger" on an age characteristic or by differentiating individuals based upon a contrived characteristic, such as "high" or "low" on a fictitious test. Standardized procedures that differentiate such abilities are the Meaning Insight, Contrast Sensitivity, or Relational Ability tasks. During these tasks, individuals are presented with binary-choice questions. The procedure proceeds in three stages. First they input their initial response. Then they are shown their partner's

response. Finally, they are given the opportunity to change or stay with their initial response. The setting is designed so that participants believe that their partner disagrees with their initial choice the majority of the time. The extent to which they stay with their initial response is the variable of interest.

The SES was constructed to test the predictions derived by the formalized derivations of SCT that use graph theory to construct predicted expectation advantages (E) between interacting partners that ultimately impact one's influence (P(s)) (see Equation 1). Those with a high expectation advantage tend to have high observed P(s) scores. Conversely, those who the theory predicts to have low expectation advantages tend to change their initial responses to match that of their perceived high ability partner. The equation below represents the probability of an actor staying with their own response, P(s), rather than changing to match that of their partner. P(s) is also influenced by two parameters, M and Q, discussed in more detail below.

$$(1)\ P(s) = m + q(E)$$

M and Q Parameters. Despite the impressive amount of standardization that the SES provides, in reality, no two settings are precisely equal. SCT implements two parameters designed to isolate the situation-specific nature of different settings; these parameters are M and Q shown in the equation above. M is the baseline tendency for someone to reject influence attempts. Q captures the effect that a particular expectation advantage has on the rejection of influence attempts (as well as any other systematic differences between conditions). Since each experiment takes place within a cultural, locational and temporal

context, each study contains unique biases. Studies can differ according to participant gender, number of status manipulations given, or completion incentive… etc. To account for these differences, the M and the Q parameters can be thought of as variables with observed values that get tweaked to account for experiment-specific idiosyncrasies and other systematic effects present in one particular experimental setting and/or condition (Berger et al. 1977). As such, these parameters tend to be estimated from the data[7] and differ from study to study.

In looking at the progression of the M and Q parameters, we see a move from situation-specific empirical estimation towards theoretical benchmarks. This is good because the ultimate goal of theoretical development is to replace such situation-specific constants with best predicted estimates. As tests of theoretical premises cumulate, this should give rise to an expanding theoretical scope. In other words, the goal is to approximate not just the baseline tendency for a particular subject to reject influence under certain conditions in a unique experiment, but the baseline tendency for anyone under any conditions to reject influence attempts. The more parameters that are theoretically derived, the closer we come to reaching such an ideal. Luckily, researchers can capitalize on SCT's long history of standardized experiments. Standardization of processes gives us confidence that computing an average M and Q derived from all applicable SCT studies should approximate the true value of M and Q for any other given study. In time, as the number of standardized studies becomes even larger, it will become increasingly likely that the average M and Q taken from these studies is equal to the first moment in the distribution of M and Q.

[7] In terms of the linear model, M and Q correspond to the slope and intercept of the model, respectively.

Kalkhoff and Thye's (2006) meta-analysis was the first step towards such a goal. They pooled SCT studies and identified conditions under which M and Q vary. They show that protocol variations, number of trials, sample size and exclusion rates all affect the observed M and Q. In addition to identifying these deviations, they calculate an average M and Q and encourage future researchers to "use these new estimates as a benchmark to assess the (ir)regularity of their own results" (Kalkhoff and Thye 2006, p. 221). Such work is primarily useful because researchers can now measure the degree to which a particular experimental setting deviates from the SES. It can even help to pinpoint if this deviation comes from an experiment-wide peculiarity in the tendency to reject influence (M) or if these differences are due to a particular condition (Q).

Novel Status Manipulations. Another consequence of Kalkhoff and Thye (2006) is that researchers who are interested in implementing novel status manipulations can now compare the strength of these manipulations to an established norm. While the SES is the foundation of SCT research, as research broadens the scope of the theory, novel status manipulations are increasingly utilized.

While the SES traditionally focused on dyadic relationships, researchers are also interested in status effects in groups sized larger than two interactants. Testing SCT predictions in a different setting dictated an alteration of the SES to accommodate these interests. Researchers adopted changes to the computer mediated binary-choice programs, measured new dependent variables, such as group consensus on item ranking during the Lost on the Moon task, and implemented novel status manipulations.

Of particular interest to the current paper is the novel status manipulation of handedness (e.g., Lovaglia et al 1998). During this manipulation, participants are told that the hand that they dominantly use is associated with either high or low task ability. Researchers have used this manipulation rather than the standard tasks described earlier for many reasons. Handedness, as opposed to something like Contrast Sensitivity ability, is more innate. Individuals have been aware of and have used their handedness for their entire lives. It is more likely that they would identify with handedness as a status characteristic than an ability that they just discovered they had or did not have. This likely increases how much an individual identifies with the manipulation and thus affects its believability (Lovaglia et al. 1998). Another consideration has been the length of standard manipulations. If the outcome variable is equally long and similar in format, participants could easily become fatigued. This could result in end effects on the dependent variable.

Knowing that even subtle changes to the SES can produce significantly different outcomes (Troyer 2001), it seems likely that changing the manipulation will cause deviations from the normative values typically obtained from the SES. When using a novel manipulation, we must ensure two things: (i) that the manipulation produces the relational differences predicted between high and low status conditions; and (ii) that the observed stay response proportions for each condition agree with the proportions predicted SCT models. The following analysis presents these results.

Study

The goal of this research is to illustrate how to use Kalkhoff and Thye's (2006) study to test the strength of a novel status manipulation. First, we present the results of a dyadic interaction that invoked status differences with the handedness manipulation. We then test two models against the data. The first model uses M and Q that are derived from the data, while the second model uses the mean M and Q taken from 26 studies compiled by Kalkhoff and Thye (2006). Finally, we present the findings of a Monte Carlo simulation that reconciles the different findings from each model.

Method

We ran a completely randomized design that used handedness as a status manipulation. Participants were told that their dominant hand was associated with either high or low task ability. They were led to believe that they were interacting with another who possessed the opposite level of competence. The dependent variable was stay response scores on the Contrast Sensitivity task.

Data

Participants were excluded from our analyses for one of the following reasons: (i) suspicion that their partner in the task was not real (7 participants) and (ii) disbelief that handedness contributed to an increased/decreased task ability (2 participants). This left 32

analyzable cases (16 per cell)[8]. Of these cases 18 were female and 14 male. All participants reported their race as white in the pre-study[9].

Findings

Table A.1. Mean and Median P(s) for Observed Data.

Condition	Observed (n=16)	
	Mean	Median
High Status	.711	.743
Low Status	.419	.425

Relational Status Differences. The first step in assessing the effectiveness of the handedness manipulation was to compare P(s) scores between high and low status individuals. Table A.1 shows the mean and median observed stay responses for high and low states of the handedness characteristic. Findings from the Mann-Whitney U test of difference in medians shows that those who were told that their handedness was associated with low task ability deferred to their partner more than those who believed that their handedness resulted in high task competence ($U = 28.00$, $p < .001$). This means that the handedness manipulation was successful in invoking the desired status differentials between high and low states of the characteristic.

[8] Studies using the standardized experimental setting typically eliminate about 15% of cases due to scope violations. Our exclusion rate of 28% is rather high, but still within the range of 0-50% observed by Dippong (2012).

[9] During the actual study, two people categorized themselves as White-Hispanic and Multi-Ethnic, respectively. We kept these cases in the analysis because the results with and without them were qualitatively similar. $\bar{x}$ of C2 (n=14) = .70 (.12).

Table A.2. Observed Cell Counts of Stay Responses and Trials Influenced.

Condition	Defer to Partner	Stay Response
High Status (C2)	92	228
Low Status (C3)	186	134

Table A.3. Predicted Cell Counts and Chi-squared Components: Data-driven.

Condition	Predicted Counts		Residuals		x^2 Components	
	Defer to Partner	Stay Response	Defer to Partner	Stay Response	Defer to Partner	Stay Response
High Status	92.31	227.69	-0.31	0.31	0.0011	0.0004
Low Status	184.92	133.08	0.08	-0.08	0.0004	-0.0006

x^2= 0.0016

Theoretical Status Differences. In addition to establishing relational differences, another goal of SCT researchers is to confirm that their observed scores fall within a theoretically predicted range. Table A.2 presents the observed cell counts of stay responses and trials in which the participant deferred to the partner for each condition. Table A.3 presents the predicted cell counts using the M and Q estimated from the data, the residuals between the observed cell counts and the expected cell counts, and the Chi-squared components. The results indicate that the model fits the data (x^2(2) = 0.0016, p = 0.9992). Observed stay responses for high and low status conditions agreed with predictions.

Table A.4. Predicted Cell Counts and Chi-squared Components: Meta-analysis.

Condition	Predicted Counts		Residuals		x^2 Components	
	Defer to Partner	Stay Response	Defer to Partner	Stay Response	Defer to Partner	Stay Response
High Status	87.71	232.29	4.29	-4.29	0.21	0.08
Low Status	149.18	168.82	35.82	-35.82	8.60	7.60

x^2= 16.49

Until recently, this analysis (coupled with the significant relational difference between high and low status conditions) would have been sufficient to conclude that the handedness manipulation was successful in invoking status differences. However, due to the methodological improvements made by Kalkhoff and Thye (2006), it is now also plausible to test the observed stay responses against a predicted model that uses meta-analysis estimates of M and Q. Table A.4 produces these results. Fitting the observed data against a model that uses M and Q derived from the average of all previously observed Ms and Qs produces a model that does not fit the data (x^2(2) = 16.49, p < 0.001). While those in the high status condition deferred slightly more than expected, the largest deviation from the model appears in the low status condition. Here, the theory predicts that subjects would defer far less than was observed.

Results from these models tell us two different things. One the one hand, there is evidence that suggests that the handedness manipulation successfully invoked status differences. We observed significant between-condition median differences. Additionally, when controlling for situation-specific idiosyncrasies by estimating M and Q from the data, our observed model fit predictions. However, on the other hand, when deriving P(s) scores by using the benchmark values of M and Q suggested by Kalkhoff and Thye (2006), we find that our empirical findings do not square with predictions. The

low status manipulation is stronger than theoretically predicted. Such contradictory results prove problematic.

Simulation Study. The following demonstrates how we reconciled the opposing findings detailed above. Kalkhoff and Thye (2006) report that M and Q are normally distributed and they report values for 26 studies. Using the data from this analysis, we calculated the mean and variance of reported Ms (m=.62, σ^2= .06) and Qs (m=.10, σ^2 =.05). We then simulated a Monte Carlo distribution of 20,000 observations of predicted M and Q values (see Appendix J for the R code). From these M and Q values, we derived predicted stay response scores for both a high and low specific status characteristics with a path length of 3 and 4 (see Appendix K for calculations).

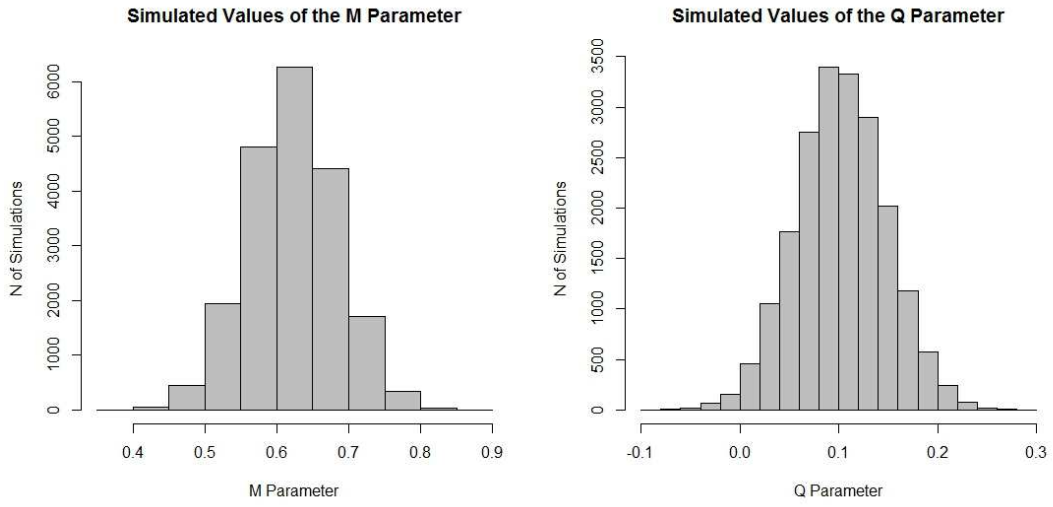


Figure A.1. Simulated Values of the M and Q Parameters.

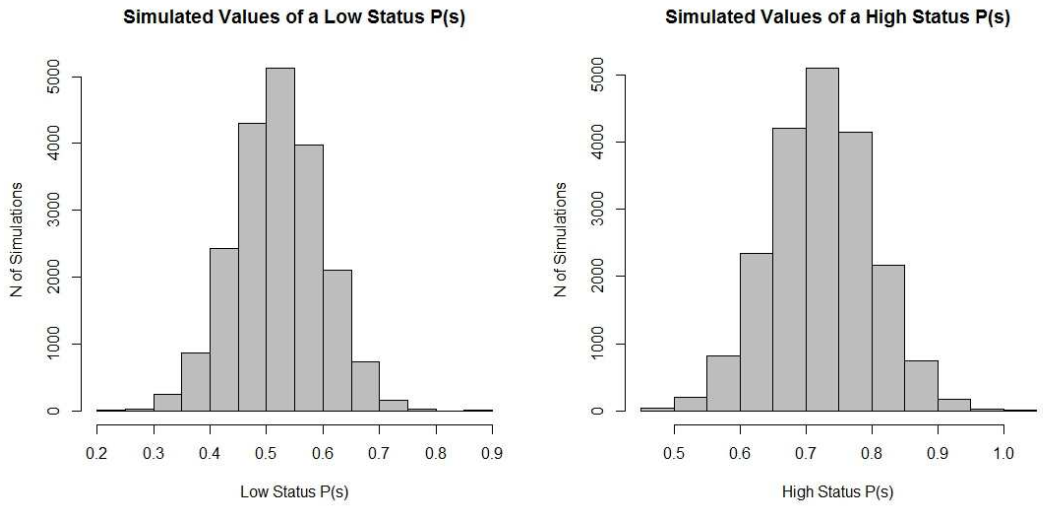


Figure A.2. Simulated P(s) Scores for a High and Low Status Actor.

Figure A.1 shows the distribution of predicted M and Q values. Values within the 95% confidence interval range between .50 and .74 for M and from .01 to .19 for Q. Of the studies cited in the meta-analysis, 96% report M values within the range observed and 92% fall within the predicted range for Q. The handedness manipulation presented earlier in this paper also is within this range. Assuming a directly relevant specific SC with a path length of 3 and 4, predicted P(s) scores for a high status actor fall within a 95% confidence interval of .57 to .87 and .37 to .67 for a low status actor. Figure A.2 produces the results from the entire simulated distribution. These results are important because they recognize and illustrate that M and Q are variables with observed values in each study. Using the means on M and Q to estimate model fit may lead to the wrong conclusion, as above, because that doesn't acknowledge that any given study may yield observed values of M and Q that are within a relatively large, and yet reasonable, range.

Discussion

Drawing on methodological advancements made by Kalkhoff and Thye (2006), we provide further refinement of the M and Q estimation process. When implementing a novel status manipulation, we suggest that researchers should meet two requirements: (i) that status differences are present between conditions; and (ii) that the predicted stay response scores fall within a predicted range. Following these guidelines will help to ensure that novel manipulations have their intended outcomes. It will also help to preserve the standardization for which the SCT research tradition is known.

APPENDIX B: STUDY ONE: SCRIPTS

Low Status, Costs

We are investigating the way people come together to form work teams. Today's work environments are intensely social. Most people no longer perform their jobs in isolation, such as on assembly lines. Rather, we usually work more closely with other people who may be different from us in many respects. Sometimes these differences make working together difficult; tempers flare and feelings get hurt.

A group has to work together to be successful. But group members normally have different levels of ability. So as the social interaction of group work becomes more intense, the ability to get along with others and coordinate the work of diverse group members becomes especially difficult.

One factor that we are studying here today is whether group members are right-handed or left-handed. We want to learn more about how these differences affect work group performance.

Today you will participate in an intense work group. Each of you will be assigned a job in a mock mental health treatment center. All of your fellow group members will be

women. Your interactions with the other group members will be observed, video-recorded and coded by trained sociologists.

Before you enter the work group, we would like to tell you more about what previous studies about right-handed and left-handed workers show.

Previous research has shown that left-handed workers are much better at coordinating group work and getting things done in groups. In general, left-handers make better leaders. Social scientists are not exactly sure why left-handers are better at these tasks than right-handers, although certain psychological processes having to do with the left and right brain are thought to cause this effect.

In contrast, right-handers are too inflexible and likely to make poor decisions. Right-handers may be so rigid that they cannot lead a group effectively, though they sometimes make good followers. For this reason, right-handers in positions of leadership are often resented by other team members. This further hampers the ability of right-handers to work in groups. In addition, right-handers are often too subservient.

In general, left-handers do much better in the kind of intense work groups that you will participate in today.

Researchers have developed several tests to determine who will be especially good at working in the kind close-knit groups that you will join later today. You will take one of these tests before joining your group.

Relative pay is another factor in how well groups perform. When a group feels that some members are overpaid, intense resentment builds up and group performance suffers. To try to optimize group performance we use your score on the aptitude tests to place you in a job. The different jobs available and how much each position pays is shown below. Your pay is determined by whether you are right-handed or left-handed and your score on the aptitude tests.

Ordinarily, everyone would get paid the same amount of money for participation. However, due to the tough economic times, we had to cut pay. We cut pay based on which job you hold during the group work. We do this because your job tells us how high you scored on the aptitude test and whether or not you are right-handed or left-handed. The job that requires the highest aptitude will not receive a pay cut. All other jobs will receive pay cuts as detailed below:

APTITUDE	POSITION	PAY CUT	PAY
Left-Handers with Superior Aptitude	Psychotherapist	-$0	$38.00
Left-Handers with Moderate Aptitude and Right-Handers with Superior Aptitude	Technician	-$24.50	$13.50
Left-Handers with Poor Aptitude and Right-Handers with Moderate and Poor Aptitudes	Orderly	-$28.00	$10.00

As you can see from the chart, your pay cut is determined by whether you are right-handed or left-handed, and by your score on the aptitude tests.

On rare occasions, right-handers with very high aptitude score are made Psychotherapist. However, we try to avoid this because of the extreme negative reactions that other group members sometimes have toward a right-handed leader. Right-handed Technicians are also resented but not as much as right-handed Psychotherapists.

Each group has one Psychotherapist, two Technicians, and three Orderlies. If you are chosen to be the Psychotherapist, you will receive the smallest pay cut of $0. Technicians will receive a pay cut of $24.50 and Orderlies will receive the largest pay cut of $28.

The job you get and your pay cut today is based on your aptitude for performing the socially complex task your group faces. We estimate your aptitude from test scores and whether you are left-handed or right-handed.

To make sure you have a good idea of the job you will be doing and how much of a pay cut you will receive, the Compensation Table will be shown again. Please study it carefully.

You will now take the aptitude test. The test that you will take is scientifically designed and nationally standardized. That means that after you finish the test, the computer will display a chart that shows you how you compare to other women in the United States.

You can also see how high you rank in comparison with the people you will work with. Group performance should be at its peak when group members have jobs that are in line with their aptitude. When you meet the other members of your group, your aptitude scores and those of other group members will be posted for all to see and compare.

We have one last thing before you take the test. Sometimes people try to cheat on the aptitude tests to get a better job or more pay. We take cheating very seriously. Cheaters will be caught and punished according to the University of South Carolina's honor code.

It is surprisingly easy to catch cheaters. We caught the last group because right-handers had scored abnormally high on the tests. We looked at the patterns of answers given by these right-handers and immediately suspected cheating.

Low Status, Rewards

We are investigating the way people come together to form work teams. Today's work environments are intensely social. Most people no longer perform their jobs in isolation, such as on assembly lines. Rather, we usually work more closely with other people who may be different from us in many respects. Sometimes these differences make working together difficult; tempers flare and feelings get hurt.

A group has to work together to be successful. But group members normally have different levels of ability. So as the social interaction of group work becomes more

intense, the ability to get along with others and coordinate the work of diverse group members becomes especially difficult.

One factor that we are studying here today is whether group members are right-handed or left-handed. We want to learn more about how these differences affect work group performance.

Today you will participate in an intense work group. Each of you will be assigned a job in a mock mental health treatment center. All of your fellow group members will be women. Your interactions with the other group members will be observed, video-recorded and coded by trained sociologists.

Before you enter the work group, we would like to tell you more about what previous studies about right-handed and left-handed workers show.

Previous research has shown that left-handed workers are much better at coordinating group work and getting things done in groups. In general, left-handers make better leaders. Social scientists are not exactly sure why left-handers are better at these tasks than right-handers, although certain psychological processes having to do with the left and right brain are thought to cause this effect.

In contrast, right-handers are too inflexible and likely to make poor decisions. Right-handers may be so rigid that they cannot lead a group effectively, though they sometimes

make good followers. For this reason, right-handers in positions of leadership are often resented by other team members. This further hampers the ability of right-handers to work in groups. In addition, right-handers are often too subservient.

In general, left-handers do much better in the kind of intense work groups that you will participate in today.

Researchers have developed several tests to determine who will be especially good at working in the kind close-knit groups that you will join later today. You will take one of these tests before joining your group.

Relative pay is another factor in how well groups perform. When a group feels that some members are overpaid, intense resentment builds up and group performance suffers. To try to optimize group performance we use your score on the aptitude tests to place you in a job. The different jobs available and how much each position pays is shown below. Your pay is determined by whether you are right-handed or left-handed and your score on the aptitude tests.

APTITUDE	POSITION	PAY
Left-Handers with Superior Aptitude	Psychotherapist	$38.00
Left-Handers with Moderate Aptitude and Right-Handers with Superior Aptitude	Technician	$13.50
Left-Handers with Poor Aptitude and Right-Handers with Moderate and Poor Aptitudes	Orderly	$10.00

As you can see from the chart, your pay is determined by whether you are right-handed or left-handed, and by your score on the aptitude tests.

On rare occasions, right-handers with very high aptitude score are made Psychotherapist. However, we try to avoid this because of the extreme negative reactions that other group members sometimes have toward a right-handed leader. Right-handed Technicians are also resented but not as much as right-handed Psychotherapists.

Each group has one Psychotherapist, two Technicians, and three Orderlies. If you are chosen to be the Psychotherapist, you will be paid the highest amount of $38. Technicians will be paid $13.50 and Orderlies will receive the lowest pay of $10.

The job you get and your pay today is based on your aptitude for performing the socially complex task your group faces. We estimate your aptitude from test scores and whether you are left-handed or right-handed.

To make sure you have a good idea of the job you will be doing and how much of a pay you will receive, the Compensation Table will be shown again. Please study it carefully.

You will now take the aptitude test. The test that you will take is scientifically designed and nationally standardized. That means that after you finish the test, the computer will display a chart that shows you how you compare to other women in the United States.

You can also see how high you rank in comparison with the people you will work with. Group performance should be at its peak when group members have jobs that are in line with their aptitude. When you meet the other members of your group, your aptitude scores and those of other group members will be posted for all to see and compare.

We have one last thing before you take the test. As a further bonus, the names of high scoring left-handers will be forwarded to the career center for special recognition. This kind of recognition is beneficial for job resumes and graduate school applications.

High scoring right-handers will receive an introductory pamphlet from the career center. While this pamphlet can be beneficial, much further effort will be required on the part of right-handers to really make a difference.

High Status, Costs

We are investigating the way people come together to form work teams. Today's work environments are intensely social. Most people no longer perform their jobs in isolation, such as on assembly lines. Rather, we usually work more closely with other people who may be different from us in many respects. Sometimes these differences make working together difficult; tempers flare and feelings get hurt.

A group has to work together to be successful. But group members normally have different levels of ability. So as the social interaction of group work becomes more intense, the ability to get along with others and coordinate the work of diverse group members becomes especially difficult.

One factor that we are studying here today is whether group members are right-handed or left-handed. We want to learn more about how these differences affect work group performance.

Today you will participate in an intense work group. Each of you will be assigned a job in a mock mental health treatment center. All of your fellow group members will be women. Your interactions with the other group members will be observed, video-recorded and coded by trained sociologists.

Before you enter the work group, we would like to tell you more about what previous studies about right-handed and left-handed workers show.

Previous research has shown that right-handed workers are much better at coordinating group work and getting things done in groups. In general, right-handers make better leaders. Social scientists are not exactly sure why right-handers are better at these tasks than left-handers, although certain psychological processes having to do with the left and right brain are thought to cause this effect.

In contrast, left-handers are too inflexible and likely to make poor decisions. Left-handers may be so rigid that they cannot lead a group effectively, though they sometimes make good followers. For this reason, left-handers in positions of leadership are often resented by other team members. This further hampers the ability of left-handers to work in groups. In addition, left-handers are often too subservient.

In general, right-handers do much better in the kind of intense work groups that you will participate in today.

Researchers have developed several tests to determine who will be especially good at working in the kind close-knit groups that you will join later today. You will take one of these tests before joining your group.

Relative pay is another factor in how well groups perform. When a group feels that some members are overpaid, intense resentment builds up and group performance suffers. To try to optimize group performance we use your score on the aptitude tests to place you in a job. The different jobs available and how much each position pays is shown below. Your pay is determined by whether you are right-handed or left-handed and your score on the aptitude tests.

Ordinarily, everyone would get paid the same amount of money for participation. However, due to the tough economic times, we had to cut pay. We cut pay based on which job you hold during the group work. We do this because your job tells us how high you scored on the aptitude test and whether or not you are right-handed or left-handed. The job that requires the highest aptitude will not receive a pay cut. All other jobs will receive pay cuts as detailed below:

APTITUDE	POSITION	PAY CUT	PAY
Right-Handers with Superior Aptitude	Psychotherapist	-$0	$38.00
Right-Handers with Moderate Aptitude and Left-Handers with Superior Aptitude	Technician	-$24.50	$13.50
Right-Handers with Poor Aptitude and Left-Handers with Moderate and Poor Aptitudes	Orderly	-$28.00	$10.00

As you can see from the chart, your pay cut is determined by whether you are right-handed or left-handed, and by your score on the aptitude tests.

On rare occasions, left-handers with very high aptitude score are made Psychotherapist. However, we try to avoid this because of the extreme negative reactions that other group members sometimes have toward a left-handed leader. Left-handed Technicians are also resented but not as much as left-handed Psychotherapists.

Each group has one Psychotherapist, two Technicians, and three Orderlies. If you are chosen to be the Psychotherapist, you will receive the smallest pay cut of $0. Technicians will receive a pay cut of $24.50 and Orderlies will receive the largest pay cut of $28.

The job you get and your pay cut today is based on your aptitude for performing the socially complex task your group faces. We estimate your aptitude from test scores and whether you are left-handed or right-handed.

To make sure you have a good idea of the job you will be doing and how much of a pay cut you will receive, the Compensation Table will be shown again. Please study it carefully.

You will now take the aptitude test. The test that you will take is scientifically designed and nationally standardized. That means that after you finish the test, the computer will display a chart that shows you how you compare to other women in the United States.

You can also see how high you rank in comparison with the people you will work with. Group performance should be at its peak when group members have jobs that are in line

with their aptitude. When you meet the other members of your group, your aptitude scores and those of other group members will be posted for all to see and compare.

We have one last thing before you take the test. Sometimes people try to cheat on the aptitude tests to get a better job or more pay. We take cheating very seriously. Cheaters will be caught and punished according to the University of South Carolina's honor code.

It is surprisingly easy to catch cheaters. We caught the last group because left-handers had scored abnormally high on the tests. We looked at the patterns of answers given by these left-handers and immediately suspected cheating.

High Status, Rewards

We are investigating the way people come together to form work teams. Today's work environments are intensely social. Most people no longer perform their jobs in isolation, such as on assembly lines. Rather, we usually work more closely with other people who may be different from us in many respects. Sometimes these differences make working together difficult; tempers flare and feelings get hurt.

A group has to work together to be successful. But group members normally have different levels of ability. So as the social interaction of group work becomes more intense, the ability to get along with others and coordinate the work of diverse group members becomes especially difficult.

One factor that we are studying here today is whether group members are right-handed or left-handed. We want to learn more about how these differences affect work group performance.

Today you will participate in an intense work group. Each of you will be assigned a job in a mock mental health treatment center. All of your fellow group members will be women. Your interactions with the other group members will be observed, video-recorded and coded by trained sociologists.

Before you enter the work group, we would like to tell you more about what previous studies about right-handed and left-handed workers show.

Previous research has shown that right-handed workers are much better at coordinating group work and getting things done in groups. In general, right-handers make better leaders. Social scientists are not exactly sure why right-handers are better at these tasks than left-handers, although certain psychological processes having to do with the left and right brain are thought to cause this effect.

In contrast, left-handers are too inflexible and likely to make poor decisions. Left-handers may be so rigid that they cannot lead a group effectively, though they sometimes make good followers. For this reason, left-handers in positions of leadership are often resented by other team members. This further hampers the ability of left-handers to work in groups. In addition, left-handers are often too subservient.

In general, right-handers do much better in the kind of intense work groups that you will participate in today.

Researchers have developed several tests to determine who will be especially good at working in the kind close-knit groups that you will join later today. You will take one of these tests before joining your group.

Relative pay is another factor in how well groups perform. When a group feels that some members are overpaid, intense resentment builds up and group performance suffers. To try to optimize group performance we use your score on the aptitude tests to place you in a job. The different jobs available and how much each position pays is shown below. Your pay is determined by whether you are right-handed or left-handed and your score on the aptitude tests.

APTITUDE	POSITION	PAY
Right-Handers with Superior Aptitude	Psychotherapist	$38.00
Right-Handers with Moderate Aptitude and Left-Handers with Superior Aptitude	Technician	$13.50
Right-Handers with Poor Aptitude and Left-Handers with Moderate and Poor Aptitudes	Orderly	$10.00

As you can see from the chart, your pay is determined by whether you are right-handed or left-handed, and by your score on the aptitude tests.

On rare occasions, left-handers with very high aptitude score are made Psychotherapist. However, we try to avoid this because of the extreme negative reactions that other group members sometimes have toward a left-handed leader. Left-handed Technicians are also resented but not as much as left-handed Psychotherapists.

Each group has one Psychotherapist, two Technicians, and three Orderlies. If you are chosen to be the Psychotherapist, you will be paid the highest amount of $38. Technicians will be paid $13.50 and Orderlies will receive the lowest pay of $10.

The job you get and your pay today is based on your aptitude for performing the socially complex task your group faces. We estimate your aptitude from test scores and whether you are left-handed or right-handed.

To make sure you have a good idea of the job you will be doing and how much of a pay you will receive, the Compensation Table will be shown again. Please study it carefully.

You will now take the aptitude test. The test that you will take is scientifically designed and nationally standardized. That means that after you finish the test, the computer will display a chart that shows you how you compare to other women in the United States.

You can also see how high you rank in comparison with the people you will work with. Group performance should be at its peak when group members have jobs that are in line with their aptitude. When you meet the other members of your group, your aptitude scores and those of other group members will be posted for all to see and compare.

We have one last thing before you take the test. As a further bonus, the names of high scoring right-handers will be forwarded to the career center for special recognition. This kind of recognition is beneficial for job resumes and graduate school applications.

High scoring left-handers will receive an introductory pamphlet from the career center. While this pamphlet can be beneficial, much further effort will be required on the part of left-handers to really make a difference.

APPENDIX C: EFFORT, TEST IMPORTANCE, AND MOTIVATION QUESTIONNAIRE

Student Opinion Scale

Please think about the test that you just completed. Mark the answer that best represents how you feel about each of the statements below.

1. Doing well on this test was important to me.

Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

2. I engaged in good effort throughout this test.

Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

3. I am not curious about how I did on this test relative to others.

Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

4. I am not concerned about the score I receive on this test.

Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

5. This was an important test to me.

Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

6. I gave my best effort on this test.

Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

7. While taking this test, I could have worked harder on it.

Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

8. I would like to know how well I did on this test.

Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

9. I did not give this test my full attention while completing it.

Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

10. While taking this test, I was able to persist to completion of the task.

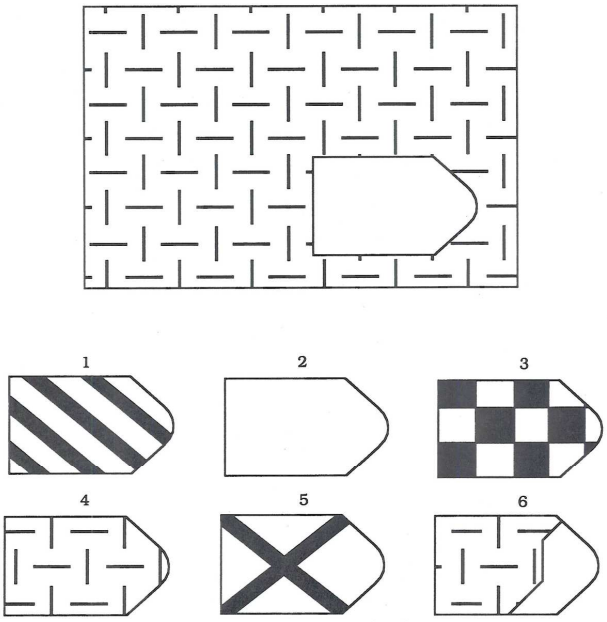
Strongly Disagree Neutral Strongly Agree

1----------2----------3----------4----------5----------6----------7

APPENDIX D: EXAMPLE OF A RAVEN'S TEST QUESTION

SET A

A1



APPENDIX E: POWER ANALYSIS TO DETERMINE SAMPLE SIZE

All figures are taken from Lovaglia et al. (1998), studies 1-3. Let

$$\emptyset^1 = \sqrt{\frac{\sum \propto j^2}{a\sigma^2 x}}$$

$$\propto = \mu_j - M$$

$$a = 2.$$

Using Feldt power charts, sample size is listed for power of .5, .7, and .9 at both the .05 and .01 levels of significance. We analyzed all three of Lovaglia et al.'s (1998) studies because they all manipulate status and measure intelligence (see Tables E.1-E.4). However, we focus our attention on Study 3 because it most closely resembles the current study (both use only women). Feldt power analysis shows that we need from 14-47 participants per cell to expect to have enough power to detect differences between conditions. Taking into consideration that the magnitude of our manipulation is less than that of Lovaglia et al. (1998), we chose to run 38 participants in each condition.

Study One

Table E.1. Descriptive Statistics of Lovaglia et al. (1998): Study One.

	Mean	SD
Condition One	55.63	3.03
Condition Two	53.91	3.60
Study Means	54.77	3.32

55.63 – 54.77 = .86

53.91 – 54.77 = -.86

$$\emptyset^1 = \sqrt{\frac{\sum(.86^2)+(-.86^2)}{2(3.32^2)}} \quad \emptyset^1 = \sqrt{\frac{\sum(.7396)+(.7396)}{2(10.989)}} \quad \emptyset^1 = \sqrt{\frac{1.4792}{21.978}} \quad \emptyset^1 = \sqrt{.0673}$$

$$\emptyset^1 = .2594$$

Table E.2. Sample Size Per Cell as Determined by Feldt Power Charts.

		Power threshold		
		.5	**.7**	**.9**
Significance	**.05**	33	52	87
level	**.01**	55	78	>100

Study Two

Table E.3. Descriptive Statistics of Lovaglia et al. (1998): Study Two.

	Mean	SD
Condition One	56.20	2.88
Condition Two	54.00	3.58
Study Means	55.10	3.23

55.63 – 54.77 = .86

53.91 – 54.77 = -.86

$$\emptyset^1 = \sqrt{\frac{\sum(1.1^2)+(-1.1^2)}{2(3.23^2)}} \quad \emptyset^1 = \sqrt{\frac{\sum(1.21)+(1.21)}{2(10.4329)}} \quad \emptyset^1 = \sqrt{\frac{2.42}{20.8658}} \quad \emptyset^1 = \sqrt{.1160}$$

$$\emptyset^1 = .3406$$

Table E.4. Sample Size Per Cell as Determined by Feldt Power Charts.

		Power threshold		
		.5	**.7**	**.9**
Significance	**.05**	18	28	46
level	**.01**	31	44	65

Study Three

Table E.5. Descriptive Statistics of Lovaglia et al. (1998): Study Three.

	Mean	SD
Condition One	54.95	2.93
Condition Two	52.35	3.39
Study Means	53.65	3.16

54.95 - 53.65 = 1.3

52.35 - 53.65 = -1.3

$$\emptyset^1 = \sqrt{\frac{\sum(1.3^2)+(-1.3^2)}{2(3.16^2)}} \quad \emptyset^1 = \sqrt{\frac{\sum(1.69)+(1.69)}{2(9.99)}} \quad \emptyset^1 = \sqrt{\frac{3.38}{19.98}} \quad \emptyset^1 = \sqrt{.1692}$$

$$\emptyset^1 = .4113$$

Table E.6. Sample Size Per Cell as Determined by Feldt Power Charts.

		Power threshold		
		.5	**.7**	**.9**
Significance	**.05**	14	19	33
level	**.01**	23	31	47

APPENDIX F: F-TEST OF NESTED MODELS CALCULATIONS: STUDY ONE

Calculation of the F statistic to compare the full model with the reduced model.

$$F = \frac{\frac{SSE\ of\ reduced\ model - SSE\ of\ full\ model}{\#\ of\ constrained\ parameters}}{MSE\ of\ full\ model}$$

$$F = \frac{\frac{1996.074 - 1990.727}{2}}{14.531}$$

$$F = .184$$

Calculation of the df for the F statistic.

$$df = (k, (n - p\ of\ full\ model))$$

$$df = (2, (144 - 7))$$

$$df = (2, 137)$$

Final result.

$$F_{(2,137)} = .184, p = .832.$$

APPENDIX G: IDENTITY QUESTIONNAIRE

Post Study Questionnaire

Instructions: Under each statement, circle the number that best describes how you feel.

Use the scale below to respond to each statement.

1. I see myself as a University of South Carolina student.

1----------2----------3----------4----------5----------6----------7----------8----------9

2. I am pleased to be a University of South Carolina student.

1----------2----------3----------4----------5----------6----------7----------8----------9

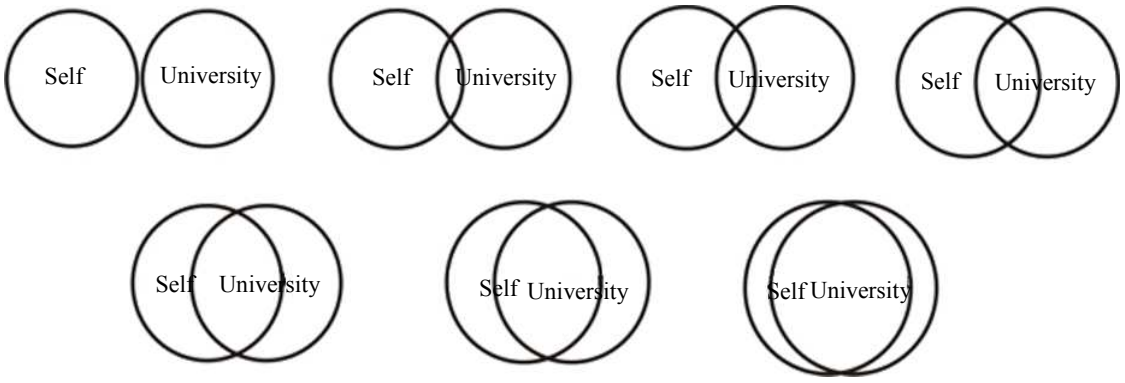
3. I feel strong ties with other University of South Carolina students.

1----------2----------3----------4----------5----------6----------7----------8----------9

4. I identify with other University of South Carolina students.

1----------2----------3----------4----------5----------6----------7----------8----------9

5. Which of the following diagrams best represents how you perceive your relationship to your university. Please circle the picture which best describes your relationship.



6. I think that I did well on the test.

1----------2----------3----------4----------5----------6----------7----------8----------9

7. I expect to do well at the group task.

1----------2----------3----------4----------5----------6----------7----------8----------9

8. I expect to get along with my group.

1----------2----------3----------4----------5----------6----------7----------8----------9

9. I expect that my group will approve of my test score.

1----------2----------3----------4----------5----------6----------7----------8----------9

APPENDIX H: STUDY TWO: SCRIPTS

In-Group Costs

We are investigating the way people come together to form work teams. Today's work environments are intensely social. Most people no longer perform their jobs in isolation, such as on assembly lines. Rather, we usually work more closely with other people who may be different from us in many respects. Sometimes these differences make working together difficult; tempers flare and feelings get hurt.

A group has to work together to be successful. But group members normally have different levels of ability. So as the social interaction of group work becomes more intense, the ability to get along with others and coordinate the work of diverse group members becomes especially difficult.

One factor that we are studying here today is college affiliation. We want to learn more about how people perform in groups together when they attend either the same or different universities.

Today you will participate in an intense work group. Each of you will be assigned a job in a mock mental health treatment center. All of your fellow group members will be

women. Your interactions with the other group members will be observed, video-recorded and coded by trained sociologists.

This is what you can expect from today's study:

1) You will be assigned to work with either University of South Carolina students or Clemson University students;

2) You will be told how your pay for the study will be calculated;

3) You will take a groupwork aptitude test; and

4) You will participate in a group task with other women from either the University of South Carolina or Clemson University.

As a reminder, please make sure to fill out the questions on the sheet in front of you when prompted by the computer.

Please answer the following questions using the choices below.

I attend college at:
The University of South Carolina
Clemson University

Please answer the following questions using the scale below.

How much do you like being a student at the University of South Carolina?

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

In 20 characters or more, please explain why you chose to attend the University of South Carolina over Clemson University.

According to research, people who attend the same university are likely to be more similar to one another than they might expect. These similarities include several dimensions, such as worldview, values, attitudes and behaviors.

This means that those who also attend the University of South Carolina may be more similar to you than those who attend Clemson University.

We do not know all of the reasons why those who attend the same college tend to be more similar to one another than those who attend different colleges, but we do know that this similarity exists and has been shown in many other research studies. One of the goals of today's study is to find out more about this.

I see myself as a University of South Carolina student.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I am pleased to be a University of South Carolina student.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I feel strong ties with other University of South Carolina students.

1----------2----------3----------4----------5----------6----------7----------8----------9

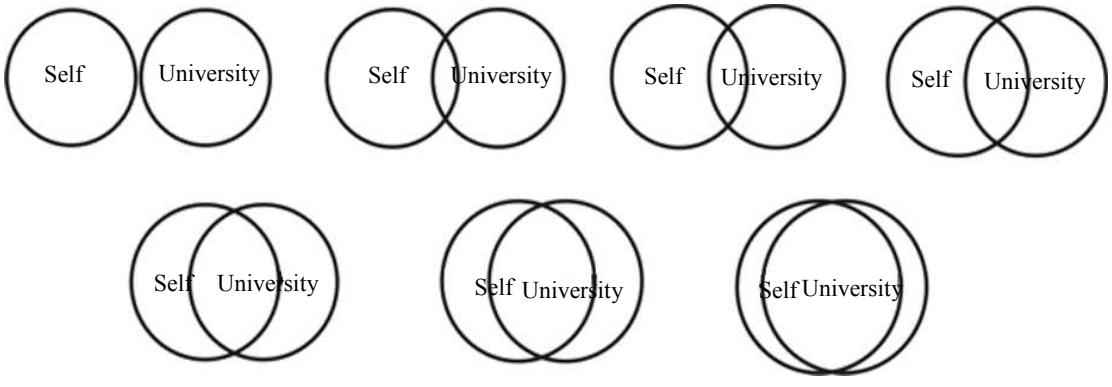
Not at all Very Much

I identify with other University of South Carolina students.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

Which of the following diagrams best represents how you perceive your relationship to your university. Please circle the picture which best describes your relationship.



One of the goals of today's research is to examine how similarities in college affiliation affects work group performance. For this reason, in the second part of the study, you will be assigned to work with students from the University of South Carolina.

Please answer Question #2 on your sheet.

We would like to tell you more about what previous studies about the students at the University of South Carolina and Clemson University show. Previous research has shown that Clemson University students are much better at coordinating group work and getting things done in groups. This means that, in general, Clemson University students make better group leaders.

Social scientists are not exactly sure why Clemson University students are better at group work than the University of South Carolina students. It could be because Clemson University is ranked almost 50 spots higher than the University of South Carolina in US News and World Report's national college rankings. Clemson University students might also be better at group work because they have a record of hiring faculty from more prestigious universities, such as Harvard University. Additionally, Clemson University offers courses that are geared specifically at improving their students' ability to participate in groups.

In contrast to Clemson University students, the University of South Carolina students are too inflexible during group work and are likely to make poor decisions as leaders. The

University of South Carolina students may be so rigid that they cannot lead a group effectively, though they sometimes make good followers. Because of the University of South Carolina's students' low ability to work in groups, we have found that during group studies run in our laboratory, University of South Carolina students that are in positions of leadership are often resented by the others in the group. This further hampers the ability of the group to succeed. In addition, the University of South Carolina students are often too subservient during group work.

In general, Clemson University students do much better in the kind of intense work groups that you will participate in today. Clemson University seems to prepare students better for group work.

Researchers have developed several tests to determine who will be especially good at working in the kind of close-knit groups that you will join later today. You will take one of these tests before joining University of South Carolina students for the group work.

Please answer Question #3 on your worksheet.

Relative pay is another factor in how well groups perform. When a group feels that some members are overpaid, intense resentment builds up and group performance suffers. To try to optimize group performance, we ran a previous study to determine the best way to pay you for your participation in this study. In this previous study, we asked University of South Carolina students or Clemson University students to rate pay charts. We asked

them to select the chart that they thought should be used to calculate your pay for the study.

University of South Carolina students decided that your pay should be determined by your college affiliation and by your score on the aptitude test. The different jobs available and how much each position pays is shown below.

Again, your pay is determined by whether you are a University of South Carolina student or a Clemson University student and your score on the aptitude test. Students at the University of South Carolina selected the following pay chart for you.

Ordinarily, everyone would get paid the same amount of money for participation. However, due to the tough economic times, University of South Carolina students decided to cut pay. They cut pay based on which job you hold during the group work. They did this because your job is based on how high you scored on the aptitude test and whether or not you attend the University of South Carolina or Clemson University. The job that requires the highest aptitude will not receive a pay cut. All other jobs will receive pay cuts as detailed below:

APTITUDE	POSITION	PAY CUT	PAY
Clemson University students with Superior Aptitude	Psychotherapist	-$0	$38.00
Clemson University students with Moderate Aptitude and University of South Carolina students with Superior Aptitude	Technician	-$24.50	$13.50
Clemson University students with Poor Aptitude and University of South Carolina students with Moderate and Poor Aptitudes	Orderly	-$28.00	$10.00

As you can see from the chart, University of South Carolina students thought that your pay cut should be determined by what college you attend and by your score on the aptitude tests.

According to University of South Carolina students, on rare occasions, University of South Carolina students with very high aptitude score can be made Psychotherapist. However, University of South Carolina students wanted to try to avoid this because of the extreme negative reactions that other University of South Carolina students sometimes have toward a University of South Carolina group leader. They thought that University of South Carolina students with the Technician job would also be resented but not as much as University of South Carolina students with the Psychotherapist job.

Each group has one Psychotherapist, two Technicians, and three Orderlies. If you are chosen to be the Psychotherapist, University of South Carolina students determined that you will receive the smallest pay cut of $0. Technicians will receive a pay cut of $24.50 and Orderlies will receive the largest pay cut of $28.

The job you get and your pay cut today is based on what college you go to and your aptitude for performing the socially complex task your group faces. University of South Carolina students decided that we will calculate your pay based on your aptitude from test scores and whether you attend the University of South Carolina or Clemson University.

To make sure you have a good idea of the job you will be doing and how much of a pay cut you will receive, the Compensation Table will be shown again. Please study it carefully and answer Question #4 on your worksheet.

You will now take the aptitude test. The test that you will take is scientifically designed and nationally standardized. That means that after you finish the test, the computer will display a chart that shows you how you compare to other women in the United States.

After you complete the aptitude test, you will be working with students from the University of South Carolina. You will also be able to see how high you rank in comparison with University of South Carolina students. Group performance should be at its peak when group members have jobs that are in line with their aptitude. When you meet the other members of the work group, your aptitude scores and those of other group members will be posted for all to see and compare.

We have one last thing before you take the test. Sometimes people try to cheat on the aptitude tests to get a better job or more pay. We take cheating very seriously. Cheaters will be caught and punished according to the University of South Carolina's honor code.

It is surprisingly easy to catch cheaters. We caught the last group because University of South Carolina students had scored abnormally high on the tests. We looked at the patterns of answers given by these University of South Carolina students and immediately suspected cheating.

I think that I did well on the test.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I expect to do well at the group task.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I expect to get along with my group.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I expect that my group will approve of my test score.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

Out-Group Costs

We are investigating the way people come together to form work teams. Today's work environments are intensely social. Most people no longer perform their jobs in isolation, such as on assembly lines. Rather, we usually work more closely with other people who may be different from us in many respects. Sometimes these differences make working together difficult; tempers flare and feelings get hurt.

A group has to work together to be successful. But group members normally have different levels of ability. So as the social interaction of group work becomes more intense, the ability to get along with others and coordinate the work of diverse group members becomes especially difficult.

One factor that we are studying here today is college affiliation. We want to learn more about how people perform in groups together when they attend either the same or different universities.

Today you will participate in an intense work group. Each of you will be assigned a job in a mock mental health treatment center. All of your fellow group members will be women. Your interactions with the other group members will be observed, video-recorded and coded by trained sociologists.

This is what you can expect from today's study:

1) You will be assigned to work with either University of South Carolina students or Clemson University students;

2) You will be told how your pay for the study will be calculated;

3) You will take a groupwork aptitude test; and

4) You will participate in a group task with other women from either the University of South Carolina or Clemson University.

As a reminder, please make sure to fill out the questions on the sheet in front of you when prompted by the computer.

Please answer the following questions using the choices below.

I attend college at:
The University of South Carolina
Clemson University

Please answer the following questions using the scale below.

How much do you like being a student at the University of South Carolina?

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

In 20 characters or more, please explain why you chose to attend the University of South Carolina over Clemson University.

According to research, people who attend the same university are likely to be more similar to one another than they might expect. These similarities include several dimensions, such as worldview, values, attitudes and behaviors.

This means that those who also attend the University of South Carolina may be more similar to you than those who attend Clemson University.

We do not know all of the reasons why those who attend the same college tend to be more similar to one another than those who attend different colleges, but we do know that this similarity exists and has been shown in many other research studies. One of the goals of today's study is to find out more about this.

I see myself as a University of South Carolina student.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I am pleased to be a University of South Carolina student.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I feel strong ties with other University of South Carolina students.

1----------2----------3----------4----------5----------6----------7----------8----------9

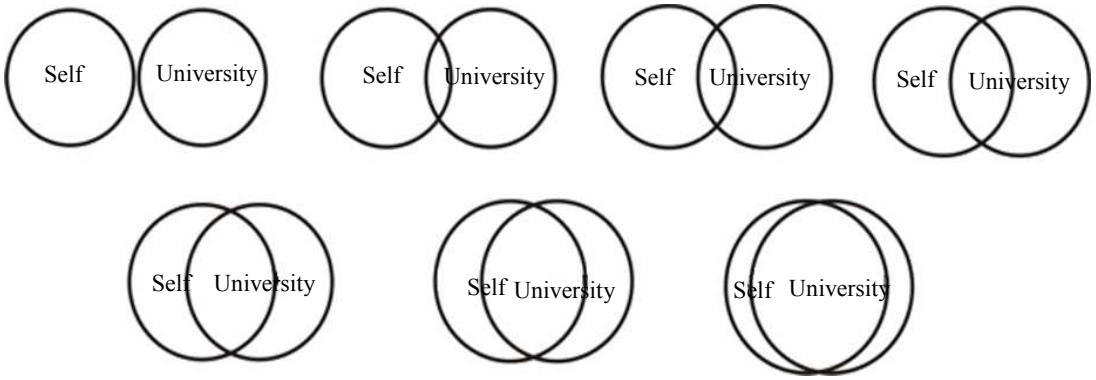
Not at all Very Much

I identify with other University of South Carolina students.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

Which of the following diagrams best represents how you perceive your relationship to your university. Please circle the picture which best describes your relationship.



One of the goals of today's research is to examine how similarities in college affiliation affects work group performance. For this reason, in the second part of the study, you will be assigned to work with students from the University of South Carolina.

Please answer Question #2 on your sheet.

We would like to tell you more about what previous studies about the students at the University of South Carolina and Clemson University show. Previous research has shown that Clemson University students are much better at coordinating group work and getting things done in groups. This means that, in general, Clemson University students make better group leaders.

Social scientists are not exactly sure why Clemson University students are better at group work than the University of South Carolina students. It could be because Clemson University is ranked almost 50 spots higher than the University of South Carolina in US News and World Report's national college rankings. Clemson University students might also be better at group work because they have a record of hiring faculty from more prestigious universities, such as Harvard University. Additionally, Clemson University offers courses that are geared specifically at improving their students' ability to participate in groups.

In contrast to Clemson University students, the University of South Carolina students are too inflexible during group work and are likely to make poor decisions as leaders. The University of South Carolina students may be so rigid that they cannot lead a group effectively, though they sometimes make good followers. Because of the University of South Carolina's students' low ability to work in groups, we have found that during group studies run in our laboratory, University of South Carolina students that are in

positions of leadership are often resented by the others in the group. This further hampers the ability of the group to succeed. In addition, the University of South Carolina students are often too subservient during group work.

In general, Clemson University students do much better in the kind of intense work groups that you will participate in today. Clemson University seems to prepare students better for group work.

Researchers have developed several tests to determine who will be especially good at working in the kind of close-knit groups that you will join later today. You will take one of these tests before joining Clemson University students for the group work.

Please answer Question #3 on your worksheet.

Relative pay is another factor in how well groups perform. When a group feels that some members are overpaid, intense resentment builds up and group performance suffers. To try to optimize group performance, we ran a previous study to determine the best way to pay you for your participation in this study. In this previous study, we asked University of South Carolina students or Clemson University students to rate pay charts. We asked them to select the chart that they thought should be used to calculate your pay for the study.

Clemson University students decided that your pay should be determined by your college affiliation and by your score on the aptitude test. The different jobs available and how much each position pays is shown below.

Again, your pay is determined by whether you are a University of South Carolina student or a Clemson University student and your score on the aptitude test. Students at Clemson University selected the following pay chart for you.

Ordinarily, everyone would get paid the same amount of money for participation. However, due to the tough economic times, University of South Carolina students decided to cut pay. They cut pay based on which job you hold during the group work. They did this because your job is based on how high you scored on the aptitude test and whether or not you attend the University of South Carolina or Clemson University. The job that requires the highest aptitude will not receive a pay cut. All other jobs will receive pay cuts as detailed below:

APTITUDE	POSITION	PAY CUT	PAY
Clemson University students with Superior Aptitude	Psychotherapist	-$0	$38.00
Clemson University students with Moderate Aptitude and University of South Carolina students with Superior Aptitude	Technician	-$24.50	$13.50
Clemson University students with Poor Aptitude and University of South Carolina students with Moderate and Poor Aptitudes	Orderly	-$28.00	$10.00

As you can see from the chart, Clemson University students thought that your pay cut should be determined by what college you attend and by your score on the aptitude tests.

According to Clemson University students, on rare occasions, University of South Carolina students with very high aptitude score can be made Psychotherapist. However, Clemson University students wanted to try to avoid this because of the extreme negative reactions that other Clemson University students sometimes have toward a University of South Carolina group leader. They thought that University of South Carolina students with the Technician job would also be resented but not as much as University of South Carolina students with the Psychotherapist job.

Each group has one Psychotherapist, two Technicians, and three Orderlies. If you are chosen to be the Psychotherapist, Clemson University students determined that you will receive the smallest pay cut of $0. Technicians will receive a pay cut of $24.50 and Orderlies will receive the largest pay cut of $28.

The job you get and your pay cut today is based on what college you go to and your aptitude for performing the socially complex task your group faces. Clemson University students decided that we will calculate your pay based on your aptitude from test scores and whether you attend the University of South Carolina or Clemson University.

To make sure you have a good idea of the job you will be doing and how much of a pay cut you will receive, the Compensation Table will be shown again. Please study it carefully and answer Question #4 on your worksheet.

You will now take the aptitude test. The test that you will take is scientifically designed and nationally standardized. That means that after you finish the test, the computer will display a chart that shows you how you compare to other women in the United States.

After you complete the aptitude test, you will be working with students from Clemson University. You will also be able to see how high you rank in comparison with Clemson University students. Group performance should be at its peak when group members have jobs that are in line with their aptitude. When you meet the other members of the work group, your aptitude scores and those of other group members will be posted for all to see and compare.

We have one last thing before you take the test. Sometimes people try to cheat on the aptitude tests to get a better job or more pay. We take cheating very seriously. Cheaters will be caught and punished according to the University of South Carolina's honor code.

It is surprisingly easy to catch cheaters. We caught the last group because University of South Carolina students had scored abnormally high on the tests. We looked at the patterns of answers given by these University of South Carolina students and immediately suspected cheating.

I think that I did well on the test.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I expect to do well at the group task.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I expect to get along with my group.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

I expect that my group will approve of my test score.

1----------2----------3----------4----------5----------6----------7----------8----------9

Not at all Very Much

APPENDIX I: F-TEST OF NESTED MODELS: STUDY TWO

Calculation of F statistic to compare Model 1 with Model 2.

$$F = \frac{\frac{SSE\ of\ reduced\ model - SSE\ of\ full\ model}{\#\ of\ constrained\ parameters}}{MSE\ of\ full\ model}$$

$$F = \frac{\frac{432.391 - 411.904}{4}}{16.476}$$

$$F = 0.311$$

Calculation of the df for the F statistic.

$$df = (k, (n - p\ of\ full\ model))$$

$$df = (4, (34 - 9))$$

$$df = (4, 25)$$

Final result.

$$F_{(4,25)} = .311, p = .868$$

Calculation of F statistic to compare Model 2 with Model 3

$$F = \frac{\frac{SSE\ of\ reduced\ model - SSE\ of\ full\ model}{\#\ of\ constrained\ parameters}}{MSE\ of\ full\ model}$$

$$F = \frac{\frac{459.48 - 432.391}{1}}{14.91}$$

$$F = 1.817$$

Calculation of the df for the F statistic.

$$df = (k, (n - p\ of\ full\ model))$$

$$df = (1, (34 - 5))$$

$$df = (1, 29)$$

Final result.

$$F_{(4,25)} = 1.817, p = 0.188$$

APPENDIX J: SIMULATION R CODE

Setting a seed ensures that results can be replicated:

```
set.seed(102)
```

The "rnorm" command takes draws off of a normal distribution based on the first two moments of the distribution. We take 20,000 draws off of the distribution of M and Q.

```
a<-rnorm(20000,mean=.621694,sd=.061694)
b<-rnorm(20000,mean=.101896,sd=.045476)
```

We then multiply Q times the expectation advantage of P over O:

```
b2<-b*.9899
```

We then "column bind" and sum M and Q(ep – eo) to generate predicted P(s) values:

```
c<-cbind(a,b2)
d<-rowSums(c)
d<-sort(d)
```

Finally, we repeat these steps for O:

```
a5<-rnorm(20000,mean=.621694,sd=.061694)
b5<-rnorm(20000,mean=.101896,sd=.045476)
b6<-b*-.9899
c5<-cbind(a,b6)
```

```
d5<-rowSums(c5)

d5<-sort(d5).
```

APPENDIX K: CALCULATIONS FOR EXPECTATION ADVANTAGES

Path lengths for P
f(3), f(4)

Path lengths for O
f(-3), f(-4)

Path weights taken from Balkwell (1991)

f (1…n) = [1-(1-f(i))… [1-(1-f(n)]

$e^{+}p$= 1- (1-f(3)) (1-f(4))

$e^{+}p$ = (1-.4056) (1-.1504)

$e^{+}p$ = (.5944) (.8496)

$e^{+}p$ = .495

$e^{-}p$= 0

e^{p} = e^{+} - e^{-}

e^{p} = .495 – 0

e^{p} - .495

Expectation advantage for P

$E = e^p - e^o$
E = .495 - -.495
E = .99

Expectation advantage for O

$E = e^o - e^p$
E = -.495 - .495
E = -.99

P(s) for P

P(s) = m + q(E)
P(s) = .6284 + .0985 (.99)
P(s) = .6284 + .097515
P(s) = .725915

P(s) for O

P(s) = m + q(E)
P(s) = .6284 + .0985 (-.99)
P(s) = .6284 - .097515
P(s) = .530885