



## CHAPTER 40

.....

# INTELLIGENCE, RACE, AND PSYCHOLOGICAL TESTING

.....

MARK ALFANO, LATASHA HOLDEN,  
AND ANDREW CONWAY

THE *Stanford Encyclopedia of Philosophy* has an entry for “Logic and Artificial Intelligence” (Thomason 2014) and “Virtue Epistemology” (Turri and Alfano 2011), but no entry for “Intelligence.” Likewise, [www.philpapers.org](http://www.philpapers.org), the largest curated database of philosophical publications in the Anglophone world, has entries for “Philosophy of Artificial Intelligence,” “Special Topics in Artificial Intelligence,” “Philosophy of Artificial Intelligence, Miscellaneous,” “The Nature of Artificial Intelligence,” “Artificial Intelligence Methodology,” “Extraterrestrial Life and Intelligence,” and “Ethics of Artificial Intelligence,” but nothing for “Intelligence,” “Human Intelligence,” or “Intelligence and Race.” As of the writing of this chapter, there were just three entries in the category “Race and IQ”: Alfano (2014), Block (1996), and Kaplan and Grønfeldt Winther (2014). A contribution on the philosophy of race and intelligence is long overdue.

That said, there is an immense amount we cannot address in this chapter. Among other things, we will discuss only superficially the ontology of race, the use of intelligence testing to justify oppression and the potential implications of this, and finally how one might mitigate the outcomes associated with racial group stigmatization in this context. This chapter has two main goals: to update philosophers on the state of the art in the scientific psychology of intelligence, and to explore more recent and relevant theoretical phenomena surrounding the measurement invariance of intelligence tests. First, we provide a brief history of the scientific psychology of intelligence. Next, we discuss the metaphysics of intelligence in light of scientific studies in psychology and neuroimaging. Finally, we turn to recent skeptical developments related to measurement invariance. These have largely focused on attributability: Where do the mechanisms and dispositions that explain people’s performance on tests of intelligence inhere—in the agent, in the local testing environment, in the culture, or in the interactions among these? After explaining what measurement invariance is in the context of intelligence testing, we will discuss stereotype threat as evidence challenging measurement invariance views of intelligence. In conclusion, we will review recent psychological theories that provide ways for combatting the pernicious and stigmatizing effects of this phenomenon.





## BRIEF HISTORY OF THE SCIENTIFIC PSYCHOLOGY OF INTELLIGENCE

---

The notion that people differ in their mental abilities is as old as philosophy itself. In *De Anima*, for example, Aristotle spoke of the faculties of the soul, which include not only perception but also imagination. References to smart or clever people can be found throughout literature as well, perhaps most notably in the works of Shakespeare, who puts these words in the mouth of Richard III: “O, ’tis a parlous boy; / Bold, quick, ingenious, forward, capable / He is all the mother’s, from top to toe.” Already in the sixteenth century, Shakespeare expresses the idea that intelligence is a trait, and moreover that it is heritable.

The scientific construct of intelligence dates to the late nineteenth and early twentieth centuries. In hindsight, it may seem surprising that it took so long for people to develop measurements of intelligence, but in the West until the nineteenth century, rationality and reason were seen by most people as divine gifts from God, so it may have seemed pointless to measure mental ability. In addition, it was not until the middle of the nineteenth century that compulsory education laws were introduced in Europe. This led to practical societal problems such as of the assessment of learning, abilities, and disabilities. Indeed, it was this need for assessment that paved the way for modern intelligence testing.

Ironically, the first scientific intelligence test was constructed primarily to detect intellectual *disability*, not to measure and compare people of normal intelligence. In 1904, the French government called upon psychologist Alfred Binet to develop a scale that could be used to identify students who were struggling in school so that they could receive alternative or supplemental instruction. Binet, in collaboration with his student Theodore Simon, developed an array of tasks thought to be representative of typical children’s abilities at various ages. They administered the tasks, now known as the Binet-Simon test, to a sample of fifty children: ten children in each of five age groups. The children in the sample were selected by their schoolteachers for being average, or representative, for their age group (Siegler 1992).

The test initially consisted of thirty tasks of increasing difficulty. The simplest tasks tested a child’s ability to follow instructions. Slightly more difficult tasks required children to repeat simple sentences and to define basic vocabulary words. Among the hardest tasks were a digit span test, which required children to recall seven digits in correct serial order, and a rhyming task, which required children to generate rhyming words given a target word (Fancher 1985). Binet and Simon administered all the tasks to the sample of children, and the score derived from the test was thought to reflect the child’s mental age. This initial standardization allowed educators to determine the extent to which a child was on par with her peers by subtracting the child’s mental age from her chronological age. For example, a child with a mental age of 6 and a chronological age of 9 would receive a score of 3, indicating that he was mentally three years behind his average peer.

While Binet and Simon were primarily interested in identifying children with learning disabilities, their methodology was quickly adapted and extended. For example, at Stanford University in 1916, Lewis Terman published a revision of their test, which he termed the Stanford-Binet test. Terman expanded the battery of tasks and adopted the intelligence quotient (IQ) rather than Binet’s difference score, an idea first proposed by German psychologist and philosopher William Stern. IQ is the ratio of a child’s mental age to her chronological age,





times 100. Thus, an average IQ is 100, and scores greater than 100 reflect higher IQ. Terman used IQ scores not only to identify children at the low end of the distribution, as Binet did, but also at the top of the distribution, as he began to study factors that lead to giftedness and genius.

While Terman and others can be credited with the first instances of group intelligence testing, the first large-scale testing was conducted with 1.7 million US soldiers during World War I. The US military, in consultation with psychologists such as Terman and Robert Yerkes, developed two tests, the Army Alpha and the Army Beta, to help categorize army recruits based on intelligence and aptitude for officer training. The Army Alpha was a text-based test that took an hour to administer. The Army Beta was a picture-based test designed for nonreaders, who made up approximately 25 percent of the recruits. The administration of intelligence tests for job placement in the military continues to this day; the modern test is known as the ASVAB (Armed Services Vocational Aptitude Battery). It was first administered in 1968 and currently consists of eight subtests: word knowledge, paragraph comprehension, mathematics knowledge, arithmetic reasoning, general science, mechanical comprehension, electronics information, and auto and shop information. The ASVAB is currently administered to over 1 million people per year.

American college admissions tests such as the SAT and ACT can also be considered intelligence tests. The history of their acronyms, however, speaks volumes. Initially, “SAT” stood for “Scholastic Aptitude Test.” In 1990, the acronym was retained, but its meaning was changed to “Scholastic Assessment Test.” Just three years later, subject-specific tests in writing, critical reading, languages, mathematics, and the sciences were introduced. These were called the SAT II battery, in which the “A” stood for “Achievement.” The old SAT became the SAT I: Reasoning Test. Finally, in 1997, the College Board, which makes the SAT, declared that the capital letters in the test’s name did not stand for anything (Applebome 1997). The ACT underwent a similar metamorphosis, standing initially for “American College Testing” but then, starting in 1996, for nothing at all.

Besides the ASVAB, SAT, and ACT, the most popular intelligence tests in use today are the WAIS (Wechsler Adult Intelligence Scale) and the WISC (Wechsler Intelligence Scale for Children), which were originally developed by psychologist David Wechsler. The WAIS and the WISC each consist of several subtests. The verbal subtests, such as vocabulary, comprehension, and general knowledge questions, are not unlike components of Binet’s original test battery. However, the nonverbal subtests, which consist of matrix reasoning, working memory, and processing speed tests, differentiate the WAIS and WISC from most other tests. These components, which Wechsler referred to as Performance IQ, are linked to a psychological construct known as *fluid reasoning*, the capacity to solve novel problems. Importantly, fluid reasoning is largely independent from prior knowledge. Furthermore, it is strongly correlated with a range of complex cognitive behaviors, such as academic achievement, problem solving, and reading comprehension.

One natural criticism of intelligence tests that employ vocabulary and general knowledge questions is that they are liable to be parochial. Vocabulary in these tests tends to be from standard English (or any language in which they are administered), but as linguists are fond of pointing out, a language is just a dialect with an army and a navy. If this is right, then intelligence tests in which standard-language vocabulary predominates may systematically privilege test takers whose culture matches the contemporary power structure, while discriminating against test takers whose culture fails to match. Perhaps the most notorious example of this comes from an SAT analogy in which runner::marathon::oarsman::reg



atta. It is not a leap to suspect that test takers who had participated in or observed a regatta would do better with this analogy than those who had not. To make a similar point, Robert Williams (1972) developed the Black Intelligence Test of Cultural Homogeneity (BITCH-100), a vocabulary test oriented toward the language and experience of urban blacks. This test included terms such as “alley apple” and “playing the dozens.” Unsurprisingly, blacks tend to outperform whites on the BITCH-100 (Matarazzo and Wiens 1977).

It was in this milieu that so-called culture-free tests of intelligence rose to prominence. The promise of such tests is that they circumvent problematic items related to parochial vocabulary and supposedly general knowledge. The most straightforward way to do this is to avoid language altogether, as in Raven’s Progressive Matrices (Raven 2003) and the Culture Fair Intelligence Test (Cattell 1949). Even these tests, however, may not be entirely culturally neutral (Aiken 1996). This is especially damning because allegedly neutral tests that obscure their own bias may be even more damaging than tests that wear their bias on their faces, as they constitute a form of what Jason Stanley calls “undermining propaganda,” which “is presented as an embodiment of certain ideals, yet is of a kind that tends to erode those very ideals” (Stanley 2015, 53).

## THE POSITIVE MANIFOLD AND THE METAPHYSICS OF INTELLIGENCE

At the same time that Binet was developing the first modern intelligence test, British psychologists were developing the statistical tools necessary to analyze the measures obtained from such tests. Sir Francis Galton, along with his student Karl Pearson, proposed the correlation coefficient, which is used to assess the degree to which two measurements are related, or covary. The Pearson product moment correlation coefficient,  $r$ , ranges from  $-1$ , which is a perfect negative correlation, to  $+1$ , which is a perfect positive correlation. Perhaps the best-replicated empirical result in the field of psychology is the *positive manifold*: the all-positive pattern of correlations that is observed when several intelligence tests, of varying format, are administered to a large sample of subjects. While the positive manifold may not seem surprising, it is important to note that, a priori, one may not have predicted such results from intelligence tests. One may have predicted instead that individuals who do well on one type of test, say vocabulary, may suffer on a different kind of test, such as mental rotation. This raises the questions: What accounts for the positive manifold? Why is it that any measure of any facet of intelligence correlates positively with any other measure of any other facet of intelligence?

One natural answer is that all measures of intelligence tap aspects of the same *general ability*. This is Spearman’s (1904) solution to the positive manifold, according to which there is a single general factor,  $g$ , of intelligence. Spearman’s model has been criticized for failing to account for the fact that some intelligence tests correlate more strongly with each other than others. For example, a verbal test of intelligence will typically correlate positively with both another verbal test of intelligence and a spatial test of intelligence, but more strongly with the former than the latter. Such patterns of clustering led Thurstone (1938) to argue for a model of intelligence that included seven primary mental abilities and no general factor. In the ensuing decades, it has become clear that both Spearman’s model and Thurstone’s models

capture part of the truth, leading to the development of higher order and bifactor models. These models have a hierarchical structure in which a general factor explains the covariance of multiple domain-specific factors (Carroll 1993).

It is important to bear in mind that these factors, whatever their exact structure, are mathematical abstractions based on interindividual differences. While it is tempting to reify them as referring to concrete intraindividual properties or processes, one must proceed with caution when identifying the grounds of intelligence. Moving from a mathematical structure to a biological or psychological process should be construed as an abductive inference to the best explanation (Harman 1965), not a straightforward identification. This is not to say that identifying the neural or cognitive mechanisms that explain intelligence is impossible, just that it is fraught with uncertainty.

Spearman notoriously identified  $g$  with “general mental energy,” a rather mysterious domain-general process. A more attractive alternative, first proposed by Godfrey Thompson in 1916, is that the positive manifold manifests itself because any battery of intelligence tests will sample processes in an overlapping manner, such that some processes will be required by a shared subset of tasks, while others will be unique to particular tasks. This idea was given a cognitive-developmental twist by van der Maas et al. (2006), who suggest that the positive manifold arises because independent cognitive processes engage in mutually beneficial interactions during cognitive development. Through a process of virtuous feedback loops, these processes eventually become correlated, resulting in the positive manifold. More recently, the idea of overlapping and mutually reinforcing processes has been corroborated by a wealth of studies in the cognitive psychology of working memory as well as the neuroimaging of executive processes and working memory. The latter studies show that regions of the brain are differentially activated by tasks that tap different facets of intelligence, with executive processes associated with both fluid intelligence and working memory capacity occurring more in the prefrontal cortex while other processes associated with crystallized intelligence occur farther back in the brain, for example, in the parietal cortex and cerebellum.

These results and the “process overlap model” that emerges from them are explored at greater length in Conway and Kovacs (2013, 2015). The basic upshot, however, should be clear: the best explanation of interindividual differences on intelligence tests and the positive manifold is that intelligent performance results from the interaction of multiple, partially overlapping processes that sometimes feed into one another both ontogenetically and synchronically. Hence, to describe someone as more or less intelligent is to say that the overlapping psychological and neurological processes that conspire to produce their behavior in the face of cognitive tasks tend to work together better or worse than the average person’s in the context in which those tasks are administered.

## RACE, INTELLIGENCE, AND THE MEASUREMENT INVARIANCE PROBLEM

For decades, some ethnic minorities, blacks, Latinos, especially, and other marginalized groups, including women and the poor, have performed worse than average on a variety of

intelligence tests. How is this phenomenon to be understood? In a lengthy article, Arthur Jensen (1969) argued that developmental interventions such as Head Start failed to attenuate the race gap in IQ. While he did not go so far as to outright assert that group differences are genetic, he has been (often still is) read as supporting such a position. His controversial thesis provoked a deluge of responses and rebuttals. In fact, Jensen's provocation may be the single most cited article in the entire field of intelligence research. A couple of decades later, Richard Herrnstein and Charles Murray (1994) in *The Bell Curve: Race and Class Structure in American Life* revived a version of Jensen's argument. Since the publication of this book, there has been a contentious dialogue, based on experimental data, surrounding the topic of innate differences in intelligence. As opposed to the pseudo-scientific theories (e.g., *On The Natural Varieties of Mankind* by Blumenbach 1775/1969) proposed about race and the inherent characteristics of personality and intelligence during the eighteenth century, Herrnstein and Murray use empirical data to support their claims. The primary argument of *The Bell Curve* is that intelligence is a cognitive construct which is dispersed in a normal distribution or "bell curve" throughout the population. Moreover, they argue that differences in *average* intelligence between different ethnic groups is not explained by a small number of exceptionally high- or low-scoring members of different groups; instead, they show that the entire distribution of intelligence scores for ethnic minority groups (especially blacks and Latinos) is shifted down from the corresponding scores for whites. The authors claim that the difference in intelligence emerges from innate differences in ability based on the notion that intelligence is approximately 70 percent heritable and 30 percent emergent from environmental factors. Herrnstein and Murray therefore suggest that there is little that minority group members can do to overcome this difference, since most of it is genetically hardwired and cannot be altered or enhanced by environmental factors.

Still more recently, Nicholas Wade (2014) once again revived these arguments in *A Troublesome Inheritance*. Wade argues that "human behavior has a genetic basis that varies from one race to another" (184), with almost the entire contemporary human population comprising just three categorically distinct races: European, East Asian, and sub-Saharan African. He goes on to claim that "national disparities in wealth arise from differences in intelligence" (189) determined in large part by race. Wade singles out "the adaptation of Jews to capitalism" (214), which, he says, accounts for the high rate at which Jews win Nobel prizes. He also argues that the genetic predisposition to "nonviolence, literacy, thrift and patience" (160) among the English upper-class eventually propagated throughout the English population, which, he suggests, accounts for the Industrial Revolution and Britain's subsequent dominance as a world empire. Of course, if race itself is not genetic, then racial differences in intelligence are a fortiori not genetically grounded. We will not address that issue here; for a summary of the main views in the philosophy of race, see Mallon (2006).

While some might be tempted to make a facile Aristotelian assumption that there is some unified property or process of individual agents in which intelligence innately inheres (e.g., something essential about their DNA), persistent differences between groups in scores on intelligence tests admit of a variety of potential explanations. Furthermore, just as we should avoid reifying *g* (the general intelligence factor) as referring to a unified, concrete process or resource in the brain, so we should be wary of inferring that someone's race straightforwardly determines his intelligence through the expression of essential genetic properties.

Indeed, as we saw in the previous section, current scientific thinking on intelligence holds that it is not unified at all, but rather emerges from the interaction of multiple, partially overlapping processes that sometimes feed into one another both ontogenetically and synchronically. Furthermore, some of these processes are *domain-specific* (that is, relative to a structure or context of learning) which means that a test of intelligence that fails to account for test takers' different experiences of domain content is liable to produce biased results (recall the "regatta" analogy). Interference with any one of the overlapping processes is likely to interfere with related processes. And when it comes to domain-general processes (e.g., working memory capacity), aspects of the testing environment that interfere with their functioning are especially likely to interfere with others.

How, then, should we interpret the well-documented differences between racial and ethnic groups on intelligence tests? In the psychological literature, this is understood as a question of *measurement invariance*. A procedure exhibits measurement invariance across groups when it measures the same construct in them. In the ideal case, it also measures the target construct with the same degree of precision and accuracy across the groups. Moreover, in order for a test to be measurement invariant, there should be no systematic bias in test performance based on group membership—measurement invariance is violated if scores on some ability are not independent of group membership. In other words, measurement invariance holds if the distribution of manifest scores of variable  $Y$  based on latent ability  $\eta$  for group  $v$  is equal to the distribution of manifest scores on variable  $Y$  based on latent ability  $\eta$ . That is, scores are invariant based on group membership alone (see Wicherts et al. 2005; Wicherts and Dolan 2010). A simple example illustrates this concept: imagine a spectrographic test that is meant to determine, based on the light a tomato reflects, whether it is ripe. Such a test could not, without adjustment, be sensibly used for both Campari and Kumato tomatoes. It also could not sensibly be used for both Kumato tomatoes in a greenhouse lighted with incandescent bulbs and Kumato tomatoes in a greenhouse lighted with fluorescent bulbs.

In much the same way, researchers have raised concerns about the measurement invariance of intelligence tests. Perhaps intelligence tests measure intelligence in whites but some other, related construct in other test takers. This would not be that surprising in light of the fact that intelligence tests were first developed for use with white populations. Or perhaps intelligence tests do measure intelligence in all Anglophone populations but are more accurate or more precise with some groups than others. Once again, the history of these tests makes this a hypothesis worth considering. Bowden et al. (2011) showed that average scores on the WAIS are slightly higher in Canada than in the United States for representative populations; this could mean that Canadians are, on average, more intelligent than Americans, but it could also mean that the test does not tap the construct of intelligence with the same accuracy in both countries.

The same holds for differences between racial and ethnic groups within the United States and elsewhere. Do the mechanisms and dispositions that explain people's performance on tests of intelligence inhere in the agent, in the local testing environment, in the culture, or in the interactions among these? This question is unavoidably laden with moral, social, and political value judgments, but there is empirical evidence that at least points toward a broadly interactionist conclusion. We will briefly review some of this evidence here; for further philosophical discussion of how best to interpret group differences in intelligence tests, see Alfano (2014) and Alfano and Skorburg (2017).

## STEREOTYPE THREAT AND MEASUREMENT INVARIANCE

Perhaps the best-known challenge to standard measurement invariance views of intelligence tests comes from the literature on *stereotype threat*, a phenomenon first investigated by Claude Steele and Joshua Aronson. They began with the idea that if you are worried that others will view your performance on a task as emblematic of your group, and your group is stigmatized as low performing or of low ability on that task, then you will experience a level of threat that people from another group might not. Such threat, in turn, is expected to lead to performance decrements. Thus, Steele and Aronson aimed to demonstrate that standard views of intelligence tests as measurements of ability were inherently flawed. Contrary to the notion that such tests measure innate ability, stereotype threat effects suggest a potential violation of measurement invariance in the form of a systematic measurement bias based on group membership. In particular, since there is a stereotype in the United States that blacks are poor students, blacks will experience a level of threat that white students do not experience on the same task. This experience, in turn, mediates performance: the more nervous you are about the inferences others might draw about you or your group based on your individual performance, the worse you are likely to do on the test. To demonstrate this, Steele and Aronson (1995) conducted an experiment with black and white undergraduates at Stanford University. Participants were randomly assigned to one of two groups. Only the first group was told that the test they were about to take was diagnostic of ability. Thus, their threat level was increased: if they performed poorly, it could reflect poorly both on them and their whole group. As predicted, black students in the first group underperformed their matched peers in the second group: merely being told that the test they were about to take was indicative of ability led to performance decrements. Based on such findings, social psychologists argue that, rather than an innate difference derived from genetics, environmental factors such as the stigma of belonging to an ethnic minority group might be responsible for much of the observed differences between groups on intelligence tests.

While the synchronic effects of stereotype threat are not fully understood (see Stricker and Ward 2004; Wei 2012; Ganley et al. 2013 for a review on the replicability and durability of the effect), one promising line of research suggests that stereotype threat interferes with working memory capacity. As we mentioned earlier, working memory capacity is highly associated with fluid *g*, a domain-general process involved in many of the overlapping processes that together constitute intelligence. For example, Schmader and Johns (2003) found that compared to controls, those under either racial or gender related stereotype threat showed significant decreases in their working memory capacity. Indeed, in this study, the effect of stereotype threat on standardized test performance on a math GRE was found to be fully mediated by decreased working memory capacity. In the same vein, Beilock and Carr (2005) have shown that in scenarios where situational pressure is induced, only individuals high in working-memory capacity “choke under pressure” on math problems that demand extensive use of working memory. Ironically, it seems that the most qualified test takers are the most susceptible to stereotype threat, in large part, because, in their case, working memory capacity resources are spent on worrying about the test and how their performance on it will be interpreted by others instead of devoting these resources to actual test

performance. Likewise, Beilock and DeCaro (2007) showed that individuals high (but not low) in working memory capacity suffered performance decrements when trying to solve multistep math problems under pressure because they reverted to unreliable heuristics, but that they enjoyed improved performance under pressure when trying to solve math problems for which a heuristic was the optimal strategy. As a whole, these studies provide insight into what might be interfering with performance for some groups taking intelligence tests compared to others: it is likely that differences in group performance could be the result of measurement bias (a violation of measurement invariance) rather than the manifestation of innate differences in ability.

Stereotype threat may also influence scores on intelligence tests diachronically by interfering not only with the *performance* of acquired skills but also with *learning* itself. In a dramatic demonstration of this, Taylor and Walton (2011) showed that black students who were tested in a threatening condition performed worse than white students in the same condition. They also showed, however, that black students who studied in a threatening condition performed even worse than other black students who studied in a nonthreatening condition. As Taylor and Walton put it, this suggests that stereotype threat generates a kind of “double jeopardy,” in which both knowledge acquisition and knowledge manifestation suffer interference. Recall that, according to contemporary theories of intelligence, independent cognitive processes often engage in mutually beneficial interactions during cognitive development. Instances of double jeopardy like those documented by Taylor and Walton (2011) are liable to derail such virtuous feedback loops during ontogenesis (i.e., individual development).

Results like these lead Wicherts et al. (2005, 705) to conclude that “bias due to stereotype threat on test performance of the minority groups is quite serious.” Whereas measured intelligence explains as much as 30 percent of the variance in numerical ability among majority test takers, it explains only 0.1 percent of the variance in numerical ability among stigmatized minorities. As they put it, “due to stereotype threat, the Numerical Ability test has become completely worthless as a measure of intelligence in the minority group” (2005, 705). This is in large part because intelligence tests are especially ineffectual in pinpointing the intelligence of highly intelligent minorities, blunting their discriminatory power.

The preceding discussion may seem unremittingly gloomy. We are therefore eager to provide some balance by pointing to interventions that have been shown to buffer against stereotype threat both synchronically and diachronically. Why do white students (especially middle- and upper-class white men) tend to fare well even in high-stakes academic tests? One theory that has recently gained purchase is that their sense of belonging in the world of elite academics shields them against the nagging concerns that mediate stereotype threat in minority students. Because they tend to be certain that they are and will be judged academically capable regardless of how they perform on any particular test, they are better able to shrug off worries about how *this* test might reflect on them or their group. By contrast, stigmatized minority students often find themselves in the precarious situation in which every mistake is taken to redound not only to their own intellectual detriment but also to the detriment of everyone else who belongs to their group. Accordingly, decreasing students’ sense of belonging in academia should make them more susceptible to stereotype threat, while increasing their sense of belonging should buffer against stereotype threat. Along these lines, Walton and Cohen (2007) showed that black (but not white) students earned higher college grades when they were given an intervention that mitigated doubts about their social belonging in college.

Beyond, but also conceptually related, to academic belongingness, self-affirmation and growth mindsets have been shown to buffer against stereotype threat and academic underachievement, respectively. An example of the former is a field study by Cohen et al. (2006), who split students in a racially mixed high school in Massachusetts into self-affirmation and other-affirmation groups. A couple of weeks into the semester, the self-affirmation group identified something they valued, then wrote for fifteen minutes about why they valued it, while the other-affirmation group identified something they did not value, then wrote for fifteen minutes about why someone else might value it. Whereas white students' grade point averages did not differ across the two conditions, black students in the self-affirmation condition ended up with grade point averages roughly .3 greater on a 4.0 scale—basically the difference between a B+ and an A-. This corresponds to a 40 percent decrease in the racial achievement gap at that school.

The growth mindset literature is primarily associated with Stanford psychologist Carol Dweck. Her basic idea is that intentional states directed at other intentional states dynamically interact with their targets, whereas intentional states directed at other types of targets typically do not. Your beliefs about your own first-order cognitive dispositions shape how those dispositions are expressed. Your beliefs about the solubility of salt do not shape how that disposition is expressed. Salt dissolves in water whether you think so or not. As Dweck and her colleagues have shown, your ability to learn partly depends on how you conceive of that ability. Dweck distinguishes two ways of conceiving of intelligence: the *entity theory*, according to which intelligence is innate and fixed, and the *incremental theory*, according to which intelligence is acquired and susceptible to improvement with effort and practice (Dweck et al. 1995). Using this distinction, she has shown that adolescents who endorse the entity theory get lower grades in school and are less interested in schoolwork (Dweck 1999), and that different regions of the brain are activated for people who endorse the entity theory when responding to mistakes: they show increased activity in the anterior frontal P3, which is associated with social comparison, and decreased activity in regions associated with the formation of new memories (Mangels et al. 2006). Together, these results suggest that people's conceptions of intelligence influence how their own intelligence is expressed.

As we saw earlier, intelligence is often conceived of as an essential and racially determined property. It stands to reason, then, that inducing people to give up the idea that intelligence is an entity might shield them from academic underachievement. In a large-scale demonstration of this idea, Paunesku et al. (2015) used both growth-mindset and sense-of-purpose interventions through online modules in the hopes of helping students persist in the face of academic challenges. Students at risk of dropping out of high school benefited from both interventions in terms of their grades and their performance in core courses. Although the interventions were least beneficial to students who were not at risk, this result strikes an encouraging note.

The current chapter aimed to survey the race and intelligence literature and update the reader on present models and interpretations from a philosophical and psychological perspective. Taken together, these findings provide compelling evidence regarding theories surrounding race and intelligence. We observe that, contrary to previous theories, intelligent behavior emerges from an interaction of abilities across several task domains, not a single unitary factor, and differences in these abilities may not result primarily from genetic inheritance. Moreover, the psychological construct working memory capacity is an important domain general factor that is related to intelligence and become disrupted when people

are put into situationally stressful or stigmatizing environments—a finding that has been observed across several aspects of social identity, including racial identity. In addition, we have reviewed empirical evidence suggesting that one's domain-specific skills are learned to a larger degree than what was once thought—providing support for the importance of environment. In sum, the current chapter highlights a more nuanced interpretation than before: intelligence cannot be summarized by just a unitary factor model or the tradeoffs between nature versus nurture and overcoming the so-called 70/30 arguments directed at explaining group and individual differences in ability. On the contrary, current psychological theory demonstrates both cognitive and social factors at work in the process of producing intelligent behavior. Moreover, as it pertains to race, there is evidence that changing the way stigmatized individuals think about themselves, their abilities, and by extension, their intelligence may mitigate stereotype threat effects and bolster academic achievement and performance.

## REFERENCES

- Aiken, L. (1996). *Assessment of Intellectual Functioning*, 2nd edition. New York: Springer.
- Alfano, M. (2014). "Stereotype Threat and Intellectual Virtue." In *Naturalizing Epistemic Virtue*, edited by O. Flanagan and A. Fairweather, 155–174. New York: Cambridge University Press.
- Alfano, M., and A. Skorburg. (forthcoming, 2017). "The Embedded and Extended Character Hypotheses." In *Philosophy of the Social Mind*, edited by J. Kiverstein. New York: Routledge.
- Applebome, P. (1997). "Insisting It's Nothing, Creator Says SAT, Not S.A.T." *The New York Times*, April 2. Accessed May 19, 2016. <http://www.nytimes.com/1997/04/02/us/insisting-it-s-nothing-creator-says-sat-not-sat.html>.
- Beilock, Sian L., and Thomas H. Carr. (2005). "When High-Powered People Fail Working Memory and "Choking under Pressure" in Math." *Psychological Science* 16 (2): 101–105.
- Beilock, S., and M. DeCaro. (2007). "From Poor Performance to Success under Stress: Working Memory, Strategy Selection, and Mathematical Problem Solving under Pressure." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33 (6): 983–998.
- Block, N. (1996). "How Heritability Misleads about Race." *Boston Review* 20 (6): 30–35.
- Blumenbach, J. [1775] (1969). *On the Natural Varieties of Mankind*. New York: Bergman.
- Bowden, S., D. Saklofske, and L. Weiss. (2011). "Invariance of the Measurement Model Underlying the Wechsler Adult Intelligence Scale-IV in the United States and Canada." *Educational and Psychological Measurement* 71 (1): 186–199.
- Carroll, J. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. New York: Cambridge University Press.
- Cattell, R. (1949). *Culture Free Intelligence Test, Scale 1, Handbook*. Champaign, IL: Institute for Personality and Ability Testing.
- Cohen, G., J. Garcia, N. Apfel, and A. Master. (2006). "Reducing the Racial Achievement Gap: A Social-Psychological Intervention." *Science* 313: 1307–1310.
- Conway, A., and K. Kovacs. (2013). "Individual Differences in Intelligence and Working Memory: A Review of Latent Variable Models." *Psychology of Learning and Motivation* 58: 233–270.
- Conway, A., and K. Kovacs. (2015). "New and Emerging Models of Intelligence." *Wiley Interdisciplinary Reviews: Cognitive Science* 6 (5): 419–426.

- Dweck, C. (1999). *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia: Taylor and Francis / Psychology Press.
- Dweck, C., C. Chiu, and Y. Hong. (1995). "Implicit Theories and Their Role in Judgments and Reactions: A World from Two Perspectives." *Psychological Inquiry* 6 (4): 267–285.
- Fancher, R. (1985). *The Intelligence Men: Makers of the IQ Controversy*. New York: Norton.
- Ganley, C., L. Mingle, A. Ryan, K. Ryan, M. Vasilyeva, and M. Perry. (2013). "An Examination of Stereotype Threat Effects on Girls' Mathematics Performance." *Developmental Psychology* 49 (10): 1886–1897.
- Harman, G. (1965). "The Inference to the Best Explanation." *Philosophical Review* 74 (1): 88–95.
- Herrnstein, R., and C. Murray. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Simon & Schuster.
- Jensen, A. (1969). "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review* 39: 1–123.
- Kaplan, J. M., and R. Grønfeldt Winther. (2014). "Realism, Antirealism, and Conventionalism about Race." *Philosophy of Science* 81 (5): 1039–1052.
- Mallon, R. (2006). "Race: Normative, Not Metaphysical or Semantic." *Ethics* 116 (3): 525–551.
- Mangels, J., B. Butterfield, J. Lamb, C. Good, and C. Dweck. (2006). "Why Do Beliefs about Intelligence Influence Learning Success? A Social Cognitive Neuroscience Model." *Social Cognition and Affective Neuroscience* 1 (2): 75–86.
- Matarazzo, J., and A. Wiens. (1977). "Black Intelligence Test of Cultural Homogeneity and Wechsler Adult Intelligence Scale Scores of Black and White Police Applicants." *Journal of Applied Psychology* 62 (1): 57–63.
- Paunesku, D., G. Walton, C. Romero, E. Smith, D. Yeager, and C. Dweck. (2015). "Mind-Set Interventions are a Scalable Treatment for Academic Underachievement." *Psychological Science* 26 (6): 784–793.
- Raven, J. (2003). "Raven Progressive Matrices." In *Handbook of Nonverbal Assessment*, edited by R. S. McCallum, 223–237. New York: Springer.
- Schmader, T., and M. Johns. (2003). "Converging Evidence That Stereotype Threat Reduces Working Memory Capacity." *Journal of Personality and Social Psychology* 85 (3): 440–452.
- Siegler, R. (1992). "The Other Alfred Binet." *Developmental Psychology* 28 (2): 179–190.
- Spearman, C. (1904). "General Intelligence, Objectively Determined and Measured." *American Journal of Psychology* 15: 201–293.
- Stanley, J. (2015). *How Propaganda Works*. Princeton, NJ: Princeton University Press.
- Steele, C. M., and J. Aronson. (1995). "Stereotype Threat and the Intellectual Test Performance of African-Americans." *Journal of Personality and Social Psychology* 69: 797–811.
- Stricker, L., and W. Ward. (2004). "Stereotype Threat, Inquiring about Test Takers' Ethnicity and Gender, and Standardized Test Performance." *Journal of Applied Social Psychology* 34 (4): 665–693.
- Taylor, V., and G. Walton. (2011). "Stereotype Threat Undermines Academic Learning." *Personality and Social Psychology Bulletin* 37 (8): 1055–1067.
- Thomason, R. (2014). "Logic and Artificial Intelligence." *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed May 19, 2016. <http://plato.stanford.edu/archives/win2014/entries/logic-ai/>.
- Thurstone, L. (1938). *Primary Mental Abilities*. Chicago: University of Chicago Press.
- Turri, J., and M. Alfano. (2011). "Virtue Epistemology." *The Stanford Encyclopedia of Philosophy*. Edited by E. Zalta. Accessed August 29, 2016. <http://plato.stanford.edu/archives/fall2016/entries/epistemology-virtue/>; <http://plato.stanford.edu/entries/epistemology-virtue/>.



- Van der Maas, H., C. Dolan, R. Grasman, J. Wicherts, H. Huizenga, and M. Raijmakers. (2006). "A Dynamic Model of General Intelligence: The Positive Manifold of Intelligence by Mutualism." *Psychological Review* 113: 842–861.
- Wade, N. (2014). *A Troublesome Inheritance*. New York: Penguin.
- Walton, G., and G. Cohen. (2007). "A Question of Belonging: Race, Social Fit, and Achievement." *Journal of Personality and Social Psychology* 92 (1): 82–96.
- Wei, T. (2012). "Sticks, Stones, Words, and Broken Bones: New Field and Lab Evidence on Stereotype Threat." *Educational Evaluation and Policy Analysis* 34 (4): 465–488.
- Wicherts, J., C. Dolan, and D. Hessen. (2005). "Stereotype Threat and Group Differences in Test Performance: A Question of Measurement Invariance." *Journal of Personality and Social Psychology* 89 (5): 696–716.
- Wicherts, J. M., and C. V. Dolan. (2010). "Measurement Invariance in Confirmatory Factor Analysis: An Illustration Using IQ Test Performance of Minorities." *Educational Measurement: Issues and Practice* 29: 39–47.
- Williams, R. (September 1972). "The BITCH-100: A Culture-Specific Test." American Psychological Association Annual Convention, Honolulu, Hawaii.

