

## Into the Blindspot

Psychological research on attitudes toward race is a relatively recent scientific endeavor, dating back less than a century. The pioneers, psychologists, and sociologists of the 1920s and 1930s undertook the first studies of Americans' attitudes toward ethnic and racial groups. Their methods were the only ones then available—they asked questions. This method of “self-report” treats people as the best authorities on their own racial attitudes. The problems with question-asking methods that were described in the last chapter notwithstanding, self-reporting was very useful in the earliest studies of prejudice. This was in part because, unlike present-day Americans, early twentieth-century Americans apparently had no qualms about openly expressing their racial and ethnic attitudes.

The story of how early researchers discovered compelling evidence of prejudices toward dozens of different groups is told in [Appendix 1](#) (“Are Americans Racist?”). That appendix further describes how research methods became increasingly sophisticated during the twentieth century, evolving toward methods that no longer relied on asking questions.

This chapter focuses on the method created by Tony in 1994—a method that gives the clearest window now available into a region of the mind that is inaccessible to question-asking methods.

## A NEW KIND OF TEST

To give you a feeling for how the new method works, we ask you to try a hands-on demonstration—quite literally hands-on because you will need to have a deck of playing cards in hand. If at all possible, please find two things—a standard deck of fifty-two playing cards and a watch or clock that displays time in seconds.

Once you have the cards and the timer, first shuffle the deck a few times and hold the cards faceup. You will be timing yourself as you perform two slightly different sorting tasks.

First you will sort the cards into two piles, with hearts and diamonds to the left, spades and clubs to the right. The second task is to sort them by putting diamonds and spades to the left, clubs and hearts to the right. Before you begin, think about these two sorting tasks and ask yourself which will be easier.

Whether or not you think that one of the tasks will be easier than the other, if you've got the cards and the timer, you're ready to start. As fast as you can, first sort the cards into two piles, hearts and diamonds to your left and spades and clubs to your right. Make a note of the number of seconds you took to do that. Next, reshuffle the deck a few times and repeat the process, but this time diamonds and spades to the left, clubs and hearts to the right.

If you were more than a few seconds faster at one task than the other, make a mental note of which of the two was faster before turning the page to learn what we expect.

Almost certainly you were faster at the first task, which allowed you to use a simple rule for the sorting—red suits left, black suits right. The second task didn't offer any such simple rule. Doing each test twice, Mahzarin and Tony averaged twenty-four seconds for the first task (red suits versus black suits) and thirty-seven seconds for the second (spades and diamonds versus hearts and clubs). Taking about 50 percent longer to do the second task is a big

difference, big enough that we could feel it as we did the sorting.

## THE IMPLICIT ASSOCIATION TEST (IAT)

Next we ask you to participate in another hands-on demonstration, for which you will again need a watch or clock that displays seconds. You will also need a pen or pencil. If you prefer not to mark this book's pages, turn ahead a few pages and photocopy the two pages of the flower-insect Implicit Association Test.

Looking at the two pages of the flower-insect test, you will see that each page has words running down the middle of two columns. To the left and right of each of the words is a small circular bubble. Your task is to place a mark in the bubble either to the left or to the right of each word, and to do that as fast as you can.

There are four sets of words:

Flower names: orchid, daffodil, lilac, rose, tulip, daisy, lily

Insect names: flea, centipede, gnat, wasp, roach, moth, weevil

Pleasant-meaning words: gentle, heaven, cheer, love, enjoy, happy, friend

Unpleasant-meaning words: damage, vomit, hurt, poison, evil, gloom, ugly

Just below are instructions and an example of the correct way to start marking the first sheet of the flower-insect test. After reading these instructions and the added suggestions just after them, you will be ready to start.

Do Sheet A first. Start at the top left and make marks as rapidly as you can. As soon as you have worked your way down the left column, continue without pause to do the right column in the same way. For each word mark the bubble to the left or right. Here are some added suggestions:

For **INSECTS** and for **words pleasant in meaning**, mark in the circle to the left. For everything else (FLOWERS and unpleasant-meaning words) mark in the circle to the right. Start at top left, go from top to bottom doing all items in order, then do the second column. At bottom right, record the elapsed time in seconds.

INSECTS or pleasant words	FLOWERS or unpleasant words	INSECTS or pleasant words	FLOWERS or unpleasant words
<input checked="" type="radio"/> WASP	<input type="radio"/>	<input type="radio"/> ROSE	<input type="radio"/>
<input type="radio"/> <b>poison</b>	<input checked="" type="radio"/>	<input type="radio"/> <b>heaven</b>	<input type="radio"/>
<input type="radio"/> TULIP	<input checked="" type="radio"/>	<input type="radio"/> ORCHID	<input type="radio"/>
<input checked="" type="radio"/> <b>enjoy</b>	<input type="radio"/>	<input type="radio"/> <b>gentle</b>	<input type="radio"/>

1. Use just a single short stroke for your marks—that will be fastest.
2. Do all the words in order. Don't skip any.
3. Definitely do not stop or backtrack to correct errors—that will make your result less accurate.
4. Timekeeping will be easiest if you start when your watch reads zero seconds—at the beginning of a minute.
5. Write your time (in seconds) to complete Sheet A at the bottom right of the sheet.
6. Then do Sheet B, which has different instructions and also different labels above the two columns. Have the changed instructions well in mind before you start Sheet B.
7. Record the number of seconds you took for Sheet B at the bottom right.

Of all the hands-on experiences in this book, this one is the most

important for starting on the path to grasp the essence of the book. Please do the flower-insect test on the next two pages now. If you'd prefer to complete this test online or on a mobile device instead, please go to [bit.ly/T8h6uD](http://bit.ly/T8h6uD).

# A

For **INSECTS** and for **words pleasant in meaning**, mark in the circle to the left. For everything else (**FLOWERS** and **unpleasant-meaning words**) mark in the circle to the right. Start at top left, go from top to bottom doing all items in order, then do the second column. At bottom right, record the elapsed time in seconds.

INSECTS or pleasant words	FLOWERS or unpleasant words	INSECTS or pleasant words	FLOWERS or unpleasant words
<input type="radio"/> ORCHID	<input type="radio"/>	<input type="radio"/> TULIP	<input type="radio"/>
<input type="radio"/> gentle	<input type="radio"/>	<input type="radio"/> enjoy	<input type="radio"/>
<input type="radio"/> ROSE	<input type="radio"/>	<input type="radio"/> WASP	<input type="radio"/>
<input type="radio"/> heaven	<input type="radio"/>	<input type="radio"/> poison	<input type="radio"/>
<input type="radio"/> FLEA	<input type="radio"/>	<input type="radio"/> ROACH	<input type="radio"/>
<input type="radio"/> damage	<input type="radio"/>	<input type="radio"/> evil	<input type="radio"/>
<input type="radio"/> DAFFODIL	<input type="radio"/>	<input type="radio"/> DAISY	<input type="radio"/>
<input type="radio"/> cheer	<input type="radio"/>	<input type="radio"/> gloom	<input type="radio"/>
<input type="radio"/> CENTIPEDE	<input type="radio"/>	<input type="radio"/> MOTH	<input type="radio"/>
<input type="radio"/> vomit	<input type="radio"/>	<input type="radio"/> ugly	<input type="radio"/>
<input type="radio"/> GNAT	<input type="radio"/>	<input type="radio"/> WEEVIL	<input type="radio"/>
<input type="radio"/> hurt	<input type="radio"/>	<input type="radio"/> happy	<input type="radio"/>
<input type="radio"/> LILAC	<input type="radio"/>	<input type="radio"/> LILY	<input type="radio"/>
<input type="radio"/> love	<input type="radio"/>	<input type="radio"/> friend	<input type="radio"/>

© Project Implicit, 2007

Number of seconds: \_\_\_\_\_

Number of errors: \_\_\_\_\_

**B**

For **FLOWERS** and for **words pleasant in meaning**, mark in the circle to the left. For everything else (**INSECTS** and **unpleasant-meaning words**) mark in the circle to the right. Start at top left, go from top to bottom doing all items in order, then do the second column. At bottom right, record the elapsed time in seconds.

FLOWERS or pleasant words	INSECTS or unpleasant words	FLOWERS or pleasant words	INSECTS or unpleasant words
<input type="radio"/> FLEA <input type="radio"/>	<input type="radio"/>	<input type="radio"/> GNAT <input type="radio"/>	<input type="radio"/>
<input type="radio"/> gentle <input type="radio"/>	<input type="radio"/>	<input type="radio"/> enjoy <input type="radio"/>	<input type="radio"/>
<input type="radio"/> ORCHID <input type="radio"/>	<input type="radio"/>	<input type="radio"/> WASP <input type="radio"/>	<input type="radio"/>
<input type="radio"/> evil <input type="radio"/>	<input type="radio"/>	<input type="radio"/> poison <input type="radio"/>	<input type="radio"/>
<input type="radio"/> ROSE <input type="radio"/>	<input type="radio"/>	<input type="radio"/> ROACH <input type="radio"/>	<input type="radio"/>
<input type="radio"/> damage <input type="radio"/>	<input type="radio"/>	<input type="radio"/> heaven <input type="radio"/>	<input type="radio"/>
<input type="radio"/> DAFFODIL <input type="radio"/>	<input type="radio"/>	<input type="radio"/> DAISY <input type="radio"/>	<input type="radio"/>
<input type="radio"/> cheer <input type="radio"/>	<input type="radio"/>	<input type="radio"/> gloom <input type="radio"/>	<input type="radio"/>
<input type="radio"/> CENTIPEDE <input type="radio"/>	<input type="radio"/>	<input type="radio"/> LILY <input type="radio"/>	<input type="radio"/>
<input type="radio"/> vomit <input type="radio"/>	<input type="radio"/>	<input type="radio"/> ugly <input type="radio"/>	<input type="radio"/>
<input type="radio"/> LILAC <input type="radio"/>	<input type="radio"/>	<input type="radio"/> MOTH <input type="radio"/>	<input type="radio"/>
<input type="radio"/> hurt <input type="radio"/>	<input type="radio"/>	<input type="radio"/> happy <input type="radio"/>	<input type="radio"/>
<input type="radio"/> TULIP <input type="radio"/>	<input type="radio"/>	<input type="radio"/> WEEVIL <input type="radio"/>	<input type="radio"/>
<input type="radio"/> love <input type="radio"/>	<input type="radio"/>	<input type="radio"/> friend <input type="radio"/>	<input type="radio"/>

© Project Implicit, 2007

Number of seconds: \_\_\_\_\_

Number of errors: \_\_\_\_\_

If you are reading this sentence without yet having done the flower-insect test, we urge you to go back a few pages and do it



before reading further. We say this only once in the book because we know that the experience of taking the tests in this book will prompt reactions of surprise and skepticism that will be well worthwhile.

After completing the flower-insect test, you may immediately know what it reveals about which part was easier, without even having to compute your results. But here's how to arrive at an exact score for your tests: For each of Sheets A and B, add your time in seconds ( $s$ ) to your number of errors ( $e$ ). Now subtract the sum of  $s + e$  for Sheet B from the sum of  $s + e$  for Sheet A.

If you were faster and had fewer errors for Sheet A than Sheet B, you have an automatic preference for insects relative to flowers. Much more likely, however, you found Sheet B to be the easier one, which reveals an automatic preference for flowers relative to insects. With the  $s + e$  scoring method, a difference of 18 or more between the two sheets shows a strong automatic preference one way or the other. A difference between 12 and 17 indicates a moderate automatic preference, and a difference between 6 and 11 reflects a slight automatic preference. If the  $s + e$  difference was less than 6, it should be considered too small to indicate either preference clearly.

A few years ago we gave this flower-insect test to a group of thirty-eight people with a doctorate in one of several academic disciplines. They took an average of sixteen seconds longer to complete Sheet A than to complete Sheet B. If you think about the fact that a runner of only moderate speed can run 100 meters in the extra time it took these PhDs to do Sheet A, you will get a sense of just how large a difference this is.

## HOW THE IAT (IMPLICIT ASSOCIATION TEST) WORKS

You have just completed the first version of what we now call an Implicit Association Test (IAT for short). Its effectiveness relies on the fact that your brain has stored years of past experiences that you cannot set aside when you do the IAT's sorting tasks. For flowers and insects, this stored mental content is most likely to help you put flowers together with pleasant words while interfering with your pairing flowers with unpleasant words. Similarly, it will likely be easier for you to connect insects with unpleasant words and harder to connect them with pleasant words. This is why Sheet B's task was probably easier for you than Sheet A's task—unless you're an entomologist or a ten-year-old boy.

In doing Sheet B, you may have had the feeling that flower names and pleasant words were not two different categories but just a single category of “good things.” Insect names and unpleasant words may similarly have felt like a single “bad things” category. Thinking of them this way may remind you of the task of sorting card suits when you were asked to sort together suits that shared a color, rather than suits that were not color-matched.<sup>1</sup>

When categories can be linked to each other via shared goodness or badness, the shared property is what psychologists call valence, or emotional value. Positive valence attracts and negative valence repels. Positive valence, which is shared by flower names and pleasant words, can function as a mental glue that bonds these two categories into one. When there is no shared valence, which is expected for most people when they try to put flower names together with unpleasant words, it is harder to find a connection between the two categories. There is no mental glue available, and this makes the IAT's sorting task on Sheet A far more challenging.

The mental glue that can allow two categories to combine into one corresponds to an ancient concept in psychology: mental

association. Hearts and diamonds have a mental association because they share the color red. For most people, flower names and pleasant-meaning words have a mental association because they share the more abstract quality of positive valence.<sup>2</sup>

In June 1994, Tony wrote the first computer program to administer an IAT, which used the same categories of flower, insect, pleasant, and unpleasant that were used in the flower-insect test you just took. As he was checking the program to make sure it worked properly, he became the first subject to try it. Here's how Tony later recalled the experience of taking that first IAT.

I had programmed the computer so that it first presented what I expected to be the easier task—giving the same response to flower names and pleasant words. As each word appeared on the screen, I was to press a key with my left hand for either insect names or unpleasant-meaning words, and a different key with my right hand for either flower names or pleasant words. Even though I had to keep track of instructions for four different categories, each with twenty-five different possible words, the task was easy—I breezed through it.

For the second task I had to press the left key for flower names or unpleasant words, and the right one for insect names or pleasant words. Within a few seconds, I could see that this was difficult. After (slowly) completing a series of fifty key presses at this task, I assumed that I would soon overcome this difficulty by practicing the task another few times. Wrong! I repeated the task several times—I did not improve at all.

Then I tried just forcing myself to respond rapidly to each word. The result was frustration. I made frequent errors—pressing the wrong key. I soon concluded that the only way I could respond accurately was to go slowly. That was the first

strong clue that this method might prove useful.

During the next several days I asked some University of Washington psychology graduate students to try the task. They too discovered the difficulty of the second task. Next, I tried it on volunteers from the university's introductory psychology courses. When I looked at their performances, it was obvious that, for almost all of them, the task that required the same key-press to flower names and unpleasant words was putting them into slow motion. It mattered only a little whether they did that task first or second.<sup>3</sup>

Those initial tryouts of the new procedure were very exciting because they suggested that the (not-yet-named) IAT could provide a useful way to measure one of psychology's long-established theoretical concepts—attitude. To psychologists, attitude has a meaning similar to its meaning in ordinary language. It refers to one's likes and dislikes such as, for example, liking flowers (a "positive" attitude) and disliking insects (a "negative" attitude). More technically, attitudes are the associations that link things (flowers and insects in this case) to positive or negative valence.

Attitudes can be expressed in many ways, including poetic verse: "Yet my heart is sweet with the memory of the first fresh jasmines [flower] that filled my hands when I was a child." Or: "They [insects] will eat up your trees. They will dig up your lawn. You can squash all you can, but they'll never be gone." The IAT captures attitudes toward flowers and insects much more prosaically, by comparing the speeds of completing two different sorting tasks.<sup>4</sup>

## THE RACE IAT

A second type of IAT soon followed—it was the first Race IAT. The change of procedure was small, replacing names of flowers and insects with the names of famous African Americans and European Americans. The new IAT was expected to reveal whether the method could measure one of our society's most significant and emotion-laden types of attitudes—the attitude toward a racial group. If it revealed a preference for White relative to Black race groups in some significant percentage of those who tried it, that might suggest the test was able to bypass the impression management phenomenon (described in [Chapter 2](#)) that is a significant source of interference in self-report methods for measuring racial attitudes. If the IAT could do that, it could be of great value in research. Even more important, it might help unveil a type of mental content that we and other social psychologists at the time were just beginning to understand—hidden biases that could not possibly be tapped by asking questions because their possessors were unaware of having them.

Perhaps you would like to try the Race IAT before reading anything about what results to expect from it. If you have access to a Web browser or mobile device, you can find the Race IAT on the Internet, at [bit.ly/TtkoCZ](http://bit.ly/TtkoCZ).<sup>5</sup> If you prefer, you can try the Race IAT in a paper-and-pencil version on the next two pages. Again you will need only a pencil or pen and a timer that can record seconds. But, before you do, please try to predict your performance. Do you think you will be:

- Faster in associating (sorting together) Black faces with pleasant words
- Faster in associating White faces with pleasant words
- Equally fast at both of these?

For the Race IAT on the next two pages, as for the previous flower-insect IAT, please try to go as fast as you can. As you complete each sheet, record the number of seconds it took to complete it and then, using the same scoring instructions as for the flower-insect IAT, compute the  $s + e$  difference between the two.

A note of caution: If you prefer not to risk discovering a result different from the one you predicted, you might want to avoid this IAT. About half of those who take this test—Tony and Mahzarin among them—obtain a result that deviates from their initial expectation.

**A**

For **pleasant words** and for **African American children's faces**, mark in the circle to the left. For everything else (**unpleasant words** and **European American children's faces**) mark in the circle to the right. Start at top left, go from top to bottom doing all items in order, then do the second column. At bottom right, record the elapsed time in seconds.

pleasant or Afr. Am. faces	unpleasant or Eur. Am. faces	pleasant or Afr. Am. faces	unpleasant or Eur. Am. faces
<input type="radio"/> disaster <input type="radio"/>	<input type="radio"/>	<input type="radio"/> agony <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> hatred <input type="radio"/>	<input type="radio"/>	<input type="radio"/> smile <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> honest <input type="radio"/>	<input type="radio"/>	<input type="radio"/> sincere <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> grief <input type="radio"/>	<input type="radio"/>	<input type="radio"/> crash <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> lucky <input type="radio"/>	<input type="radio"/>	<input type="radio"/> diamond <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> peace <input type="radio"/>	<input type="radio"/>	<input type="radio"/> sweet <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> rotten <input type="radio"/>	<input type="radio"/>	<input type="radio"/> tragedy <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>

Number of seconds: \_\_\_\_\_  
 Number of errors: \_\_\_\_\_

**B**

For unpleasant words and for African American children's faces, mark in the circle to the left. For everything else (pleasant words and European American children's faces) mark in the circle to the right. Start at top left, go from top to bottom doing all items in order, then do the second column. At bottom right, record the elapsed time in seconds.

unpleasant or Afr. Am. faces	pleasant or Eur. Am. faces	unpleasant or Afr. Am. faces	pleasant or Eur. Am. faces
<input type="radio"/> disaster <input type="radio"/>	<input type="radio"/>	<input type="radio"/> agony <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> hatred <input type="radio"/>	<input type="radio"/>	<input type="radio"/> smile <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> honest <input type="radio"/>	<input type="radio"/>	<input type="radio"/> sincere <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> grief <input type="radio"/>	<input type="radio"/>	<input type="radio"/> crash <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> lucky <input type="radio"/>	<input type="radio"/>	<input type="radio"/> diamond <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> peace <input type="radio"/>	<input type="radio"/>	<input type="radio"/> sweet <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>
<input type="radio"/> rotten <input type="radio"/>	<input type="radio"/>	<input type="radio"/> tragedy <input type="radio"/>	<input type="radio"/>
<input type="radio"/>  <input type="radio"/>	<input type="radio"/>	<input type="radio"/>  <input type="radio"/>	<input type="radio"/>

© Project Implicit, 2007

Number of seconds: \_\_\_\_\_  
 Number of errors: \_\_\_\_\_

Here is how Tony remembers his experience in taking the Race IAT for the first time:



I programmed the first Race IAT within a few months after the flower-insect IAT. It used names of famous African Americans and famous European Americans in place of flower and insect names. I tried it immediately when the program was ready. Because I had no preference (or so I thought) for one race group over the other, I expected to be as fast in sorting Black names together with pleasant words as in sorting White names together with pleasant words.

It was a rare moment of scientific joy to discover—in midperformance—that the new method could be important. It was also a moment of jarring self-insight. I immediately saw that I was very much faster in sorting names of famous White people together with pleasant words than in sorting names of famous Black people together with pleasant words. I can't say if I was more personally distressed or scientifically elated to discover something inside my head that I had no previous knowledge of. But there it was—it was as hard for me to link names of Black people and pleasant words as it had been a few months earlier to link insect names and pleasant words.

After taking that first Race IAT and repeating it several times to see if the first result would be repeated (it was), I did not see how I could avoid concluding that I had a strong automatic racial preference for White relative to Black—just as I had a strong automatic preference for flowers relative to insects.

I then asked myself what any social psychologist would: Is this something that affects my behavior in relation to African Americans whom I regularly encounter—especially students in my classes? Do I act as if I feel less positive toward them than toward White students?

The question Tony asked himself points to the deeper issues raised by the test. What exactly does an “automatic preference for White relative to Black” mean? Is it a sign of prejudice, and if so, what are the effects of that prejudice? If a person such as Tony, who genuinely believes himself not to be prejudiced, takes the test and then discovers a preference for White in himself, should we expect that he is likely to express this hidden bias in ways that could be damaging to others?

## DOES “AUTOMATIC WHITE PREFERENCE” MEAN “PREJUDICE”?

The race IAT holds up a mirror in which many see a reflection that they do not recognize. Most who take the Race IAT are faster on Sheet B (linking racial White to pleasant words) than on Sheet A (linking racial Black to pleasant words). This is the pattern that is described as showing “automatic preference for White relative to Black.”

In our own first experiences with the Race IAT, both of us were surprised to discover how much more easily we associated White than Black with pleasant. Our initial “There must be some mistake” reaction soon gave way to “Does this mean that I am prejudiced?” Since then, that same question has been directed to us many, many times by others who have taken the Race IAT and were confronted with test scores that were at odds with both their expectations and their self-perceptions.

For almost a decade after the Race IAT was created, when people asked us if a White-preference result means “I am prejudiced,” we dodged the question by saying that we didn’t yet know. We would say that the Race IAT measured “implicit prejudice” or “implicit bias,” emphasizing that we regarded these as clearly distinct from prejudice as it has generally been understood in psychology.

We had good reasons to be cautious. First of all, the IAT results that reveal automatic White preference—results based on speed of responding to words and pictures—bear little resemblance to the extremely negative racial attitudes that were expressed in self-reports from the twentieth century (these are described in [Appendix 1](#)). Results obtained with those question-asking methods have established an understanding that prejudice is an attitude that encompasses dislike, disrespect, and even hatred. Nothing about the IAT suggests that it taps such hostility.

The second good reason for our unwillingness to equate the IAT's "automatic White preference" result with "prejudice" was the unavailability of any research evidence needed to justify that conclusion. Neither we nor anyone else had done studies to determine whether those who show the highest levels of automatic White preference on the Race IAT are also those who are most likely to show racially discriminatory behavior.

But this situation has changed. Because of the rapid accumulation of research using the Race IAT in the last decade, two important findings are now established. First, we now know that automatic White preference is pervasive in American society—almost 75 percent of those who take the Race IAT on the Internet or in laboratory studies reveal automatic White preference. This is a surprisingly high figure. We (Mahzarin and Tony) thus learned that we are far from alone in having a Race IAT result that reveals that preference.<sup>6</sup>

Second, the automatic White preference expressed on the Race IAT is now established as signaling discriminatory behavior. It predicts discriminatory behavior even among research participants who earnestly (and, we believe, honestly) espouse egalitarian beliefs. That last statement may sound like a self-contradiction, but it's an empirical truth. Among research participants who describe themselves as racially egalitarian, the Race IAT has been shown, reliably and repeatedly, to predict discriminatory behavior that was observed in the research. Because this conclusion is surprising and therefore may not be easy to grasp, we take some space here to describe a few key portions of this large (and still rapidly growing) body of research evidence.

## **DO MERE ASSOCIATIONS SHOW UP IN BEHAVIOR?**

The first experiment to test whether scores on the Race IAT were related to discriminatory behavior was reported in 2001 by Allen McConnell and Jill Leibold, psychologists at Michigan State University. Their research subjects were forty-two Michigan State undergraduate volunteers. Without initially informing their research subjects of the fact, the researchers videotaped these student subjects during two brief interviews, one conducted by a White woman, the other by a Black woman. During the interviews the students were asked a series of innocuous pre-planned questions, such as “What would you change to improve psychology classes?” and “What did you think about the difficulty level of the computer task?” (The computer task was the Race IAT, which had been presented as if it was part of a separate experiment.)

The purpose of the videotaping was to assess whether strong automatic White preference shown on the Race IAT would predict acting in a friendlier fashion to the White interviewer than to the Black one. After completing both interviews, the experimenters explained the purposes of the videotaping and asked the students to give their permission to analyze the videotapes. All but one gave permission.

The videotapes of the interviews were scored by counting occurrences of nonverbal behaviors that had been found, in many previous studies, to indicate friendliness or coolness. Indicators of comfort or friendliness included smiling, speaking at greater length, laughing at a joke told by the interviewer, and making spontaneous social comments. Discomfort indicators included speech errors and speech hesitations. Another measure of comfort or discomfort was how closely the subjects positioned their rolling desk chair to each of the interviewers. Immediately after each interview, both of the interviewers also made personal assessments of how friendly and

comfortable they thought the subject had seemed during their interaction.

McConnell and Leibold found that subjects with higher levels of automatic White preference on the IAT showed less comfort and less friendliness when talking with the Black interviewer than with the White interviewer. This was an intriguing result, but it was also just one study that, by itself, was not enough to support a conclusion that the Race IAT could predict racially discriminatory behavior. But quite a few other researchers, aware of the possible importance of what the Race IAT could reveal, began to use the Race IAT in combination with measures of discriminatory behavior or judgment—aiming to see how well the Race IAT would predict.

Following are a few examples of behaviors that were found, in various studies, to be predicted by the Race IAT's measure of automatic White preference: in a simulated hiring situation, judging White job applicants more favorably than equally qualified Black applicants; emergency room and resident physicians recommending the optimal treatment—thrombolytic (blood-clot dissolving) therapy—less often for a Black patient than for a White patient who presented with the same acute cardiac symptoms; and college students being more ready to perceive anger in Black faces than in White faces.

By early 2007, thirty-two studies had been done in which the Race IAT was administered together with one or more measures of racially discriminatory behavior. These studies were among 184 in a collection that was published in 2009, using the statistical method of meta-analysis to combine all of these results for the purpose of evaluating the IAT's success in predicting a wide variety of judgments and behaviors.

The meta-analysis answered the most important question about which we had been uncertain in the first several years of the IAT's existence: It clearly showed that the Race IAT predicted racially

discriminatory behavior. A continuing stream of additional studies that have been completed since publication of the meta-analysis likewise supports that conclusion. Here are a few examples of race-relevant behaviors that were predicted by automatic White preference in these more recent studies: voting for John McCain rather than Barack Obama in the 2008 U.S. presidential election; laughing at anti-Black racial humor and rating it as funny; and doctors providing medical care that was deemed less satisfactory by their Black patients than by their White patients.<sup>7</sup>

The meta-analysis's findings can be summarized as saying that IAT scores correlated moderately with discriminatory judgments and behavior. "Correlated moderately" is a statistical term that needs elaboration for its implications to be fully clear. (Readers who are content not to immerse themselves in the technical explanation of that phrase can safely skip the next six paragraphs.)

The statistic used in almost all tests of the IAT's ability to predict behavior is the correlation coefficient—a number that can range between 0 and 1. Finding a correlation of 0 between a Race IAT measure and discriminatory behavior means that knowing a person's Race IAT score provides absolutely no information about the likelihood of that person engaging in discriminatory behavior. A perfect correlation of 1 would mean that knowledge of where a person ranked, from lowest to highest, on the Race IAT measure of automatic White preference would tell you exactly where he or she ranked, from lowest to highest, in discriminatory behavior. Among researchers, there is a conventional understanding that a correlation of .10 is small, one of .30 is medium, and one of .50 or greater is large. In saying that the meta-analysis found a moderate average correlation between Race IAT measures and measures of discriminatory behavior, we mean that the average correlation was close to the conventional "medium" value of .30. (To be precise, the meta-analysis found an average correlation of .24 between

Race IAT measures and discriminatory behavior.)

An example that has nothing to do with race discrimination may make the real-life implications of “moderate correlation” more meaningful. Suppose you are a bank manager whose job it is to decide which of the bank’s loan applicants should receive loans they have applied for. Although most borrowers will pay back at least some portion of their loan, some will not pay enough to yield a profit for the bank. Fortunately, there is some information that can help you judge the suitability of each potential borrower: their credit rating scores.

We need one more assumption to make use of this example: We shall assume that, as bank manager, you know that only half of potential borrowers are likely to repay enough to reach the bank’s profitability threshold. If loan applicants’ credit rating scores were known to correlate perfectly with their expected amount of repayment, your problem would be entirely solved. You should give loans to the 50 percent of applicants with the highest credit ratings, knowing that you would thereby be giving loans to all of those who will pay back enough to make the loan profitable—and only to those. You would thereby be maximizing the bank’s profit.

Of course, credit rating systems are not perfect—the correlation between credit rating scores and amount of loan repayment will not be the perfect value of 1. Let us assume that the credit ratings available to you are known to correlate at a medium (.30) level with the expectable amount of repayment. This tells you that if you give your loans to the 50 percent of applicants who have the highest credit rating scores you can expect 65 percent of them to repay at a level that will create profit for the bank, compared to only 35 percent of the loans being profitable if you instead loan to the 50 percent in the lower half of credit rating scores. Although this is clearly not a perfect outcome, notice how much more desirable it is than what would happen if you had no credit rating



scores at all—in that case half your loans would be profitable and half not, meaning little or no profit for the bank. A credit rating score that has a medium “predictive validity” correlation will therefore enable a substantial profit, even if not the maximum possible profit.<sup>8</sup>

This mortgage loan example tells us how to understand the meta-analysis’s average correlation between the Race IAT and discriminatory behavior. The correlation value of .24 means that, in a situation in which discrimination might be shown by 50 percent of those for whom we have Race IAT scores, we can expect that discrimination to be shown by 62 percent of those having automatic White preference scores in the top half of the overall distribution of scores, compared to 38 percent of those in the bottom half (difference = .24).

To the surprise of many, the meta-analysis found that the Race IAT predicted discriminatory judgments and behaviors significantly more effectively than did the types of question-asking measures that had long been used in studies of prejudice. Those self-report measures yielded an average validity correlation of only .12, compared to the IAT’s .24. The magnitude of this superiority of prediction by the IAT was not expected by anyone.

Perhaps we appear ready to conclude that those who show automatic White preference on the Race IAT should indeed be characterized as “prejudiced”—in the sense that they are more likely than others to engage in discriminatory behavior. However, one important aspect of the research evidence has yet to be taken into account, and it is a critical aspect. The forms of discrimination investigated in the research studies that used Race IAT measures involved no overtly racially hostile actions—no racial slurs, no statements of disrespect, and certainly no aggressive or violent actions. Recall the examples of the research behavior that we mentioned—social behaviors in interracial interviews, doctors’

treatment recommendations for a cardiac patient, and evaluations of job applicants in a hiring situation. These are not the types of negativity or hostility that are generally taken to be characteristic of “prejudice.”

This is why we answer no to the question “Does automatic White preference mean ‘prejudice’?” The Race IAT has little in common with measures of race prejudice that involve open expressions of hostility, dislike, and disrespect. Even so, the hidden race bias revealed by the Race IAT is unwelcome news to many who receive an automatic White preference result from the test, and it is probably also distressing to these same people to learn now that the Race IAT is a moderate predictor of racially discriminatory behavior. Included in those who are thus distressed are Mahzarin and Tony, who were not pleased to discover that hidden race bias was an uninvited potential mindbug, revealed to them when the IAT made it possible to look into their blindspot.<sup>9</sup>