

ANNOUNCER: The following content is provided by MIT OpenCourseWare under a Creative Commons License. Additional information about our license and MIT OpenCourseWare in general is available at ocw.mit.edu.

JEREMY WOLFE: Good afternoon. For the past couple of lectures, what I've been doing is using love and romance as a way to talk about broad issues like evolutionary psychology that could be talked about in a wide range of other-- with a wide range of other examples. Love and romance just happened to provide a particularly good set of examples for that particular topic.

I'm going to do the same thing now with attitude formation and the links between attitudes and behavior. But I'm going to switch from the love and relationships topic to the topic of racist or prejudicial behavior and attitudes.

Again, not because that's the only set of attitudes that are interesting or important, but because it makes for an interesting path through this material. So when you read the book, you'll get the topics discussed in rather more general terms, and here, I'll discuss them in the specific terms of this particular problem.

I should note at the outset, what it says about halfway down I see the first page of the handout, which is, I'm going to do my best to give an explanation about why prejudiced attitudes are easy to come by and easy-- not easy-- are readily comprehensible in terms of psychological processes that we actually know something about, but explaining these things is not the same thing as excusing them.

You can have a society that says, look, all else being equal, racial prejudice, and particularly, behaviors based on-- prejudices based on gender, race, national origin, religion, that those biases are bad, and when they lead to behavior, we want to change that behavior. That's quite separate. Not unrelated to, but it's separate from the question of how you would explain it psychologically. It's important to remember that explanation is not the same thing as excuse.

Because I don't want people going out of the lecture saying, my psych professor explained that it's really, really easy to develop prejudicial attitudes and that's OK. It's that "that's OK" part that-- well, maybe-- I also put on the handout a pair of quotes from the early days of the Civil Rights movement, one from Eisenhower saying that you can't legislate morality, and a response from Martin Luther King saying, well, maybe not, but you can legislate moral behavior.

That's really the social policy point. If you decide you don't like something that biology, psychology, or whatever is pushing people towards, you simply have to do something to make it harder for them to go where that tendency might push them.

Well, in any case, what I'm going to do is work a story about the development of prejudicial attitudes that's got four-- well, I guess they're called factors here. Four factors listed at the top. And I'll work through each of them. I hope you can see how they're tied together to make prejudice a very available option to us, why this thing happens to us with some regularity.

Well, the first factor that I list there is ethnocentrism. That's the tendency to think that your group is the best group, the "We're number one" kind of thing. If you want to come up with, let's say, an evolutionary psych argument for why this might happen, it's not-- I mean, it's kind of trivial, if you've got some notion that you want to get, oh, I don't know, your genes into the next generation, then you might as well favor the people who are more closely related to you.

So you should be more favorable to you-- to humans than to mice. You should be likely to be more favorable to people within your group, and even more so to people within your family out of this fairly straightforward application of evolutionary theory, for example.

The more interesting aspect psychologically is how easy it is to get ethnocentric effects-- or to get the "we're number one" kind of effects. Interesting evidence for this comes from these minimal attachment experiments that I put the-- I call it minimal group affiliation experiments. There's a lot of them, let me describe a couple to you.

So here's an experiment. You come into the lab, and we're going to do an assessment of your taste in abstract art. And we're going to show you a bunch of abstract pictures, and you're going to say how much you like it on a scale of, I don't know, 1 to 7 or something like that.

And then the feedback that you're going to get from this at the end is that you like the work of Paul Klee, one abstract artist, better than you like the work of Wassily Kandinsky, another abstract artist-- or it might be the other way around. I don't even remember, by the way, whether they were actually using real Klee and Kandinsky pictures. But you're going to get told that you're in the Klee group or the Kandinsky group.

This is not a group assignment where there is a lot at stake. So let us suppose that you're in the Klee group. You've been assigned to the Klee group.

In the second part of the experiment, you're playing a game. I don't quite remember what the story was. And it ends up with you being able to operate under one of two payoff rules. In one rule, you're going to give the other group \$1 and you're going to get \$2. That's one possibility. The other possibility is a rule that gets you \$3 and gives them \$4. So this is you, this is them. You've got a choice between option A and option B.

Clearly the rational choice from your vantage point is option B because you get \$3 and not \$2. But in fact, there's a bias towards picking this option. Why is that? Well, if you get \$2, you're getting more than them. Here, I'm going to get more than I would have gotten there. But these guys are going to get a whole bunch more. Why should I give those Kandinsky lovers a bunch of stuff?

Well, that's perhaps a little bit surprising, but maybe it turns out that after you look at-- you figure that these Kandinsky sorts really are a different sort of scummy kind of person who you really ought to stick it to. Well, in fact, there's no difference between these groups, the group assignment is completely random, so any such distinction that you had made in your own mind is meaningless, but maybe you didn't know that.

So we can rerun this whole experiment by saying what we-- get rid of the silly cover story. We're going to flip a coin. You're in group A. You're in group B. B, A, A, B, B, B, B, B. And look, guys, it's random. There is nothing differentiating your group from the other group. Guess what? You get the same result. I think it does get a little bit weaker, but not much.

So just the act of being in a group causes you to be inclined to favor that group. You are also inclined to think that group membership is diagnostic. Let me tell you what that means.

Let me describe a different experiment-- actually, it's the easiest way to describe this. Suppose I put up a bunch of dots, and we're above the subitizing range here, obviously. OK, quickly, how many dots are there there? Well, I don't know, you could guess. I mean-- and you'd either be-- you might get lucky and be right on, but you'd either be above or below.

So you're going to do this for a while, guess at a bunch of how many dots there are. And we are going to declare that you are-- and how do I want to write this out? You're an overestimator an underestimator. So that's the group assignment in this case.

Now the interesting thing in this experiment is that you've done it with another person. And the possibilities are that-- let's see here, how we-- this is a good way to write it. It could be two guys, both of whom are overestimators. It could be two guys, one of them an overestimator and one of them an underestimator. Similarly for females, and similarly-- they could both be underestimators.

And then the critical conditions are it could be a male and a female who are both overestimators-- both labeled as overestimators or both labeled as underestimators. And the critical condition is male overestimator, female underestimator, or the other way around. But the important point here is that a male is-- the male is identified as one, the female is identified as the other.

So after you-- so you've got people in all these various groups. I don't know, somebody can figure out how many groups there must be for the full design here, but you got people in all those groups. And now you ask a question, at the end of the whole experiment, do you think there's a sex difference, a systematic sex difference between men and women on this task? Or what do you think? What do you think the sex difference is if there's a sex difference?

All these groups, the male-male, male-male-- or all the same sex groups-- say, on average, I don't think there's a difference. Some people say I think women are better, some people think men are overestimators or whatever, but there's no systematic bias here. Nothing systematic happens here.

Here, something systematic happens. In this particular version of it, if you're the male subject-- well, it doesn't matter which subject you-- you would declare that males as a group are overestimators and females as a group are underestimators.

What's the evidence for that? One of each. Clearly not meaningful. There's no statistical reliability that you could glean from this. It could be that, I don't know, whatever other difference-- brown-haired people and blond-haired people or something are different.

But as soon as you've got-- you know that there's a group difference here because males and females are definitely different groups. As soon as you've got evidence that there's a difference across this group, you are willing to start making assumptions that that difference applies to the group as a whole. You see how that works? More or less? Somebody's nodding their head, that's encouraging.

OK. So that's this notion that-- that's factor 1, the idea that you've got-- that you're inclined to see your group as number 1. And this is a point we'll come back to, which is that you're inclined to see groups as having properties that pertain to the group and you'll jump to that quickly.

Now the tendency to think that all members-- that group identification tells you something about properties of the group, that's known as stereotyping. If I know that you are part of this group, I believe I know something about you, period. It also-- when we talk about stereotyping, we tend to think of it in negative terms. That is a fairly self-evident consequence of factor 1 mixed with factor 2.

If you are inclined to think that your group is number 1, and you're inclined to think that group identity tells you something, it follows that being a member of another group means you're a member of a group that is, at best, number 2. Ain't number 1, that's my group.

So you in this-- there's this group that I've identified, are in some other lesser group on this scale. That's fairly obvious. What's a little less obvious and is at least worth mentioning is that stereotypes are not just descriptions of things that are common to that population.

Here's the silly example. Let us consider these stereotypes-- the Asian woman stereotype. Is it part of the Asian woman-- is bipedality part of the Asian woman stereotype? No. Nobody sits around and says, oh, those Asian women have two feet. It's stupid. nobody's going to say anything like that because everybody's got two feet.

What's important in a stereotype, stereotypes are different scores, in a sense. It's not necessarily accurate once you understand they can be completely bogus, but what defines a stereotype is what somebody thinks differentiates a population from-- one group from the population as a whole.

So I put some data down here from one big study of stereotypes. Please note that these are not facts. I mean, they're facts in the sense that they're data. They're not true facts about, in this case, the German population. What they are is-- what this particular group of subjects reported believing about different groups-- in this case, the Germans. So we have-- I excerpted from a huge study the factors efficient, extremely nationalistic, scientifically-minded, and pleasure-loving.

You will note that the largest single category-- the highest value for the German population in this collection of data points, is pleasure-loving but that's not part of what would be considered the stereotype here. Because it's not higher than the population as a whole.

This particular group of subjects asserted that 82% of people in the world would be pleasure-loving and a mere 72% of the Germans would be. Therefore, pleasure-lovingness would not be considered part of that stereotype because it's not something that distinguishes this German-- this view of Germans from the view of people as a whole.

And, in control, efficient, yes. Extremely nationalistic, yes. And interestingly, something like scientifically-minded would be considered part of the stereotype even though it's not even held to be a majority-- this perception doesn't say that a majority of Germans are scientifically-minded, but more Germans are scientifically-minded according to this view than the population as a whole.

Again, I have no idea what the true data would be for the German population. I don't know if they think about science at all. But the perception is that the stereotype would include efficient, nationalistic, and scientific and not pleasure-minded because of this issue of being of-- a differential relationship to the perceived baseline to the population as a whole.

One of the factors contributing to the power of stereotyping is what's known as the outgroup homogeneity effect. So the in-group is you. You're in some group. The outgroup are other people. The outgroup homogeneity effect is the tendency to think that all of them are kind of alike. You don't think that about you in the same way.

So, I mean, let's-- from the last lecture, the nerdy high school freshman group, I would know if that was my in-group. I would know that they aren't all alike because I'm one of them and I know that I'm different from all those other nerdy high school freshmen and stuff like that. But the jocks, man, they're all alike. I don't know. It's in effect-- an ignorance effect.

Now, how does this play into bias? How does this ignorance factor play into not just thinking that all those people are alike, but to thinking less of those people? If you ask about what you know about groups that you don't interact with much, you will typically-- where do you get your information about them? You get your information from the news.

What gets you onto the news? The fact that you a good student and you love your mother? No. But if you decide to go off and commit an armed robbery or something like that, that might get you on the news.

And if you are a member of a distinctive group of some sort that's not my group, I'm going to say, hey, look, I've got a data point about, oh, we can keep picking on the Asian women. I've got a data point about Asian women. That one committed an armed robbery.

I know about Asian women now. They all commit armed robberies! Well, look, you don't do anything quite that bald and stupid, but that's the sense in which you are willing to color the entire group on the basis of whatever information you have about one member of it, that's another version of this effect.

Is likely to lead to negative assessments of the outgroup because the information you get about groups that you don't interact with is skewed towards the negative. The stuff that's going to make the news about a group is going to be typically the negative effect.

Now, that wouldn't make-- oh, so where are the strongest stereotypes in the American population? Well actually, this is a somewhat old study at this point, but when you assess-- you can use various questionnaire assessments to assess how strongly a population holds stereotypic views of another population.

The heavily stereotyped views held by an American population-- this is now about 15 years ago-- were held of the Turks, Arabs, to a lesser degree, the Japanese. Groups that the were not heavily represented in the American-- the general American mix. Less heavy stereotypes for groups with large immigrant populations in this country. Because there, you were more likely to know some people in that group, and that makes the stereotypes less firm, less strong.

Now, one of the reasons that these stereotypes matter is because along with being willing to build them quickly and easily, we are also inclined to think that the attributes that we put on other groups are causal, at least in others. This is what's known as the fundamental attribution error. Let me explain a little bit.

I sent a note to my social psych-- a couple of my social psych colleagues yesterday as I was thinking about this lecture asking, who was it who named the fundamental attribution error? Because that's a great thing to be able to do, to be able to call what you work on fundamental. It's a good-- and it turns out to be Nisbett, N-I-S-B-E-T-T, from the University of Michigan if you want to track that down.

But in any case, let me explain what this is. There are two broad ways of thinking about personality, about what it is-- we all have some notion about that, that we've got a personality. And thinking about the attributes that make up that personality can be divided into two broad categories that map onto the usual nature-nurture kind of arguments in the field.

There are trait theories that are on typically on the more nature side of it, that there are fundamental attributes of personality may be coming from a genetic origin that are just-- they you who you are because you have these traits. The alternative, from the more nurture side of things, the environmental side of it, is a situation-- too many letters-- -alist? A situationalist account that says you who you are because of where you are.

So, trait theory. You are here now because you were born smart and hardworking and studious, or you basically have consistent, over time, hardworking, studious attributes to you.

The situationist account says you're sitting here right now emitting student behavior because you're in a student environment. If we put you into-- we put you out in the farm, you would not be sitting there in the middle of the field with a notebook taking notes about the cow. That's not what you would be doing. In that situation, you'd be doing farm kind of stuff.

Like all such debates in the field, the truth is going to lie somewhere in between, there's going to be bits of both. You're not going to get any mileage out of arguing strictly one or strictly the other. What you think about this, what you think about the balance here is important for policy purposes.

For instance, why did this guy commit a crime? Now we know he committed a crime because we just convicted him of committing this crime. Why did he commit a crime? Is it because he's a criminal? He's a criminal sort of person. Fundamentally dishonest, nasty kind of person. Or is it because he was in a situation that promoted criminal behavior?

Why does that matter? Well, if you're in this mode, you're going to-- you may assign him to-- you may send him off to prison in both cases. In this case, you're likely to think of prison as a place where you put bad people to keep them out of the way for a length of time that's appropriate to whatever bad thing they did, but it's basically as a punishment.

This is the personality theory that would lead you to call your prison a correctional institution because you think that you could correct this person. If you think he's fundamentally a bad person, the trick is to make sure he can't do that anymore. If you think it's because of the situation that somehow forced him into-- or pushed him into criminal behavior, you want to fix that.

And so modern prison philosophy is neither all the way here or all the way there, but the balance is, it's really a personality theory question.

Now, the fundamental attribution error is a tendency to hold to a more trait-theoretic position when you're talking about other people than when you're talking about yourself. Why is it an error? Well, it logically can't be the case that you are the product largely of the situation and that they are a product of their invariant parts. I mean, that-- over the population as a whole isn't going to hold up.

So, that would be, why did this person-- why did this guy rob the bank? Well, he robbed the bank because he's a criminal. Why did I rob the bank? Why I robbed the bank, because I was hungry and the door was unlocked, and I didn't really rob the bank, I just picked up the money that was lying on the floor, and anyway, it was the other guy. Much more likely to give a situational--

Or why did you get a bad grade on the last test? Let's say the midterm. Suppose you got a bad grade on the midterm. Why did you get a bad grade on the midterm? Well, the story was really lame and it distracted me, and I didn't get enough sleep, and the course wasn't a big priority for me and that's why I got a bad grade.

Your TA says-- well, your TA is a good person and doesn't say this, but your TA looks at the exam and says, why did they get a bad grade on the exam? They're stupid! That would be the fundamental attribution error. Why did your TA get a bad grade on the last-- well, I didn't get enough sleep and stuff like that.

So we're more inclined to give situationalist accounts of our own behavior and more inclined to give trait accounts of other people's behavior. All right. So let's see where this has gotten us to. We're inclined to put people into groups. We're inclined to assign attributes to those groups. We're likely to assign more negative attributes to groups than we ought to because our information about groups that we don't know is being skewed in that direction.

And we're likely to attribute behavior to what we now perceive correctly or incorrectly as the traits of the group. And so you can see how you're going to end up with a story that is a pretty negative account of some outgroup or other.

Oh, by the way, there's a very interesting wrinkle on the fundamental attribution error, or at least there's certainly used to be. I do not know whether this is still the case in the population as a whole, and I certainly don't know, though I would be very interested to know, whether it's-- whether it's true at MIT.

Here is-- well, let's try this as an experiment and see how your intuition goes for this. Oh, OK, we can do the trait versus situational thing. And we're going to have-- oh, how do I want to do this? Oh, OK. So math test. This is particularly strong in mathematics.

So you take a math test, and you-- now we're talking about-- you're interpreting your own score on a math test. That score can be good or bad. As you may know, there are explanations of why the test score was good. A trait theory would be, I'm a genius. A situationalist theory would be, I'm lucky. All right, so we can actually-- so--

And if the test is bad, you can have an assessment that says, I'm dumb. Or you can have an assessment that said, oh, I don't know, the test is unfair. The interesting bit is that one gender is more likely to say this and the other gender is more likely to say this. One gender is more likely to say this and the other gender is more likely to say this. So each of these cells can be associated with a gender. So, I got a good grade on the test. I'm a genius. Who are we talking about?

STUDENTS: Men.

JEREMY WOLFE:All right.

[LAUGHTER]

And it follows that this must be the female cell. I got a bad grade on the test. I'm dumb. And that is the historical finding. This is a study-- these are studies that came out in the early days of the Women's Movement, and that's-- and I don't know about whether or not the-- it still pertains.

So the disturbing finding at the time was that males were inclined to give trait-theoretic answers for the good stuff-- I'm good, I'm bright, I'm gorgeous-- and females were likely to say, I'm lucky and it's all makeup, or something like that.

And on the other side, on the bad side, the females were likely to give trait-theoretic answers. I'm dumb and ugly and depressed and it's just terrible. And the guys were likely to say, yeah, you know, I'm brilliant, I'm gorgeous, et cetera, and the test was really unfair and my teacher hated me and stuff.

So it would be interesting to know whether that was still true, but-- well, here, we might as well take a poll. How many people think that if we actually collected data, we'd find that something like that was still true? How many think that we would find that it had gone away? All right. Well, so intuition suggests that it's still true. As I say, I have no idea whether-- if there's new data on that.

Anyway, the basic point is that with that modulation, possibly, we tend to see situational explanations of our own behavior and trait explanations of other people's behavior.

Now, let's take a look at this last factor, which I'm calling the role of ignorance in person perception, and see how that lets us-- how that can lead to what looks like a biased outcome, perhaps even if you didn't have these other factors.

These other factors-- this ignorance factor is now going to interact with these other factors to make biased outcomes quite easy to come by. So let's do a version of the classic physics joke about "Assume the horse is a sphere," stuff like that. We're going to oversimplify the situation. The issue here is who-- I gotta get my picture the same as on the handout here. But who are you going to be friends with?

Well, all right. First of all, we have to go back to the good, earnest high school discussion about making friends with people and stuff like that, and does it matter if they're wearing the latest design or whatever? And the answer, of course, is no because you shouldn't judge a-- book by its cover. Good. Nice cliché. And it is, of course, true.

Now, let us take-- in the first "Assume the horse is a sphere" oversimplification of the problem, let us assume that the set of people with whom you might be friends in the world is the set of-- let's just make it all MIT undergraduates. And we don't want to judge books by covers, so therefore, what you're going to do is set up in-depth interviews with everybody and decide who's going to be your friend on the basis of that. No, that's not going to work.

Well, all right. The other alternative is you don't want to judge a book by their cover, therefore, I won't talk to anybody ever again. That's not going to work either. So it is self-evident again-- various bits of this are self-evident, you've gotta make snap decisions on the basis of imperfect information.

It doesn't mean that you have to make your decisions on the basis of whether or not they're wearing designer whatevers, of course, but you're necessarily going to have to make a first cut through the population on the basis of essentially superficial information.

Well, what's that going to do? Let us assume-- where's my little picture here? Let us assume that in the world, in an act of massive further oversimplification, there are bad people and good people. And that your job is to divide the world into those two categories. So you're going to make an assessment. You're going to perform an act of person perception and divide the world into bad people and good people.

So we got a nice simple 2-by-2 design here. Ideally, you want everybody to be in those two populations and in those two cells, but the issue about not being able-- I mean, even if you could do in-depth interviews with everybody, it's not clear you'd never make a mistake.

But clearly, if you're just going to be basing your decisions on relatively superficial information, there are going to be errors. I labeled these type I and type II, which is actually jargon from signal detection land, but don't worry about that. It just, in this case, gives us a chance to ask the question, all right, which of these type of errors is worse? Which error--

If you had a choice about which error to make, which is the type of-- how many people, given the choice between making a-- let's be clear about this. You got a good person and you declare that good person to be bad, or you've got a bad person you declare them to be good. You got a choice between which of those errors you're going to make. How many people vote for type I? How many people vote for type II?

OK. So a type II person. Why do you want to make it just a type II error? Or why do you prefer the type II error? He doesn't want to-- good. Well, any type II person-- there's a type II person.

STUDENT: You're missing out on good people because--

JEREMY WOLFE: You're missing out on good people. That's the nice person answer. You don't want to inadvertently tar some nice good person with the label of being bad. How about a type I person? Somebody who voted-- yeah?

STUDENT: There should be plenty of good people out there, so it's better if you miss a couple of them than let a few bad people in.

JEREMY WOLFE: Yeah. There's a lot of people who could be your friends. You don't need all of them, and you want to get rid of those bad people. How come? Anybody else wanted to-- that person behind you had a--

STUDENT: It can be harmful?

JEREMY WOLFE: Yeah. I mean, it could be dangerous if you let that-- if we dichotomize this into good and really bad, nasty, dangerous people, then that intuition becomes a little clearer, that you-- whatever else you do, you want to keep these people away from you.

Now the problem-- this is a problem-- this is derived-- this is applied signal detection theory. Usually you do signal detection theory in visual perception land or something, but let's draw this picture. This is what the next page of the handout has on it, but here's where this is coming from.

Let us suppose, again, for the sake of vast oversimplification, that there are, on a scale on-- a scale of goodness-- so this is going to be bad to good. That there are only two types of people in the world. There are good people and bad people. And that you want to pick the good people and reject the bad people.

But the difficulty is that you can't successfully-- you can't see this because your information is lousy. The effect of your information being lousy is that what you see is a distribution of goodness and badness, something like that. Oh, I don't have any colored chalk. Oh well.

By the way, if you were doing this in vision land, this would be, say, one light and another light. Can you tell the difference between a dim light and a bright light or something like that? By the time it goes through your nervous system, rather than being always looking exactly like this or always looking exactly like this, sometimes the bright light looks a little dim, sometimes the dim light looks a little bright. And how do you decide which one you've seen?

So you've got a good person, you've got a person in front of you. How can you decide whether they're good or bad? Well, the best you can do is put a criterion in there somewhere. So let's just divide it. If I do that, that's OK. That means I'm now saying-- so I'm going to declare everybody on this side to be good and everybody on this side to be bad.

And OK, so I'm declaring all of these people, who are, in fact, good, to be good. The difficulty-- the sad thing-- and so which errors are these? This is the good people who I'm declaring to be bad. So the type I errors are here. These are good people who I declared to be bad. That's too bad. OK.

Here on this side are all the bad people who I declared to be bad. That's exactly what I wanted to do. But these guys-- how am I going to indicate this? Vertical lines or something. These are the type II errors. These are bad people who beat my criterion level and I said they're good.

Now, you should be able to figure out, looking at a picture like this, that there's no way to eliminate error. There's nothing that-- if this is the situation of the stimuli that I have to deal with, there's no way to eliminate error. All I can do is apportion error. So if I decide that the type II errors are the dangerous errors that I need to avoid, then I'm going to move my criterion over.

This, I think, is the second-- yeah, OK. So the second picture on the handout. If I move my criterion over so that I reduce my type II errors to just these few, let's say, well, now, the result is that I've massively increased my type I errors, I'm now declaring all these lovely people to be people I don't want to make friends with.

It's sad, but that's the way it goes, because as the gentleman back there said, yeah, look, there's a lot of people here still. I got plenty of people to be friends, and these guys will just have to deal with the fact that they're not my friend. That's just the way it is. But I'm not letting any of these mean, nasty, rotten people in except for these guys. Most of these guys I'm going to avoid.

OK, now, look what happens when you deal with an outgroup, when you deal with a group other than your own. If the argument is that part of what makes an outgroup the outgroup is the fact that you don't know enough about-- you know less about them, the result is that-- the way to express that in these signal detection terms is as an increase in the noise, an increase in the spread of these distributions.

So what that's going to end up looking like is you still have the good people and the bad people, but now these-- your perception is less accurate. Oh no, that's a very ugly blob. Oh well. Is it going to matter? No, it doesn't really matter. Well, it does sort of matter. Let me fix that. Get it symmetrical here. Something like that. OK, that'll do.

So they now just overlap more because you just don't know as much about these people. You're not as good at-- you want a trivial example of this, look. Out there, let's take wolves. There's an outgroup for you. Not wolfs, but wolves, the ones with the big, sharp-- "Grandma, what big teeth you've got" kind of wolves.

Maybe there's a good wolf out there somewhere. A nice wolf. The kind of wolf that we're supposed to have adopted back in antiquity to make into dogs eventually. A nice wolf out there. But you meet a wolf on the street and you don't know much about him, where should you draw your threshold before bringing him home to play with your six-year-old?

I mean, you're going to draw your threshold out here somewhere. I don't care if I reject the one nice wolf. It's just really risky to bring wolves home. And that's because you're just really, really ignorant about wolves. You just don't know much about them. Maybe if you knew wolves better, you'd know who the nice ones were.

All right. With human populations, it's obviously much less dramatic than that, but you don't know much about these other people, so the distributions theoretically overlap more. You still only want to make a very few errors where you let bad people in next to you, so that's going to cause you to move that threshold still further over in this direction.

Not because you don't like these people. Understand that there's no explicit bias going on here. You're just being cautious in this story. So you can get explicit bias out of those first three factors, but here, this factor has no explicit bias in it at all, just ignorance. So now you say, oh, good, I'm only letting this percentage of really bad people in.

Now obviously, there's a little problem here. Your type-- we're calling type I errors, these errors where you reject good people, great. Now you have rejected almost the entire population of this other group. You know that you're not biased in your heart of hearts because you can still say, as it-- oh yes, as it says on the handout, this little tail of the distribution is the sum of my best friends are x.

Whatever that outgroup is, saying, "Some of my best friends are white, Black, Christian, Jewish," whatever the other outgroup is that you're dealing with, this signal detection story will get you there with some people in the group who are fine because they beat your threshold, and the vast bulk of the rest of it who disappear because you're applying the same caution to an outgroup that you were applying to the group that you knew something about.

So that's how ignorance can end up being a factor in producing what looks like biased behavior. If you compare these two, you'd have to say, I'm biased against this group because these people-- a hundred of these people, I'm only letting five of them be my friend; a hundred of these people, I'm letting 60 of them be my friend. That's a biased outcome from no explicit bias.

Now this question of explicit versus-- explicit versus either no bias or implicit bias is an interesting one. You don't necessarily have a clear idea of the biases that you may have. One of the more interesting and more disturbing bits of-- oh, I didn't bring my computer to do a demo, what a pity.

There's a thing called an implicit attitude-- is it Implicit Attitude Test? Yeah, it is Implicit Attitude Test-- the IAT. If you want to try this out on yourself, go to www.prejudice.com I think is the right site. That is one site. But if that fails, go and find your way to the website of Mahzarin Banaji at Harvard, and she is one of the leading practitioners of this, and her website will link you to a place where you can try this out on yourself.

As the website will tell you, be forewarned, the results-- you may find the results of this experiment to be disturbing to you, but it's well worth trying out yourself. Now what is this experiment about? This experiment is, in effect, a version of a Stroop Interference Test. The classic Stroop Interference experiment is an experiment where what you do is you see a collection of words and your job is, whatever the word says, is just to tell me what color the ink is that the word is written in.

So if I write "cat" in red ink, you say "red." And if I write "dog" in blue ink, you say "dog." No, you say blue. Well, that's a different interference.

The problem is, that if I write "red" in blue ink, you will-- some people will simply make the mistake of saying "red," and everybody, on average, will be substantially slowed down because of an inability to suppress that response. And there'll be speed-- if I do "red" in red ink, they'll be speeded up. So if the two sources conflict with each other, you're slowed. If the two sources agree with each other, you're speeded.

OK, so here's what you do in an experiment. What you do is you show-- you tell people, I'm going to show you some words. And if they're good words, they're nice words, you push this button. And if they're nasty words, you push this button. No problem. So nice-- boink! Evil-- boink! Pain-- boink! And so on. Not very tough.

OK. Second task. I'm going to show you some pictures of people. If it's an old person, I want you to push one button. If it's a young person, I want you to push another button. So, we put me up there, and boink! Put you up there, boink! All right, that works.

Now what we do is we do mixed blocks where you're going to see words and pictures together. And I'm going to tell you, OK, now, if you see a nice word or an old face, push this button. If you see a nasty word or a young face, push this button. It's just the two tasks on top of each other. You'll be slower-- whatever we do in this regard, you'll be slower because now you have to keep two rules in mind at once.

But what's striking is that if you do nice and old, you're significantly slower than if you do nice and young. It is as if nice and old mapped to the same response, it causes a conflict that doesn't work for you, and nice and young does, as if you've got a bias in favor of young over old.

And this-- and if we ask why is this an implicit attitude test-- if we give you an explicit attitude test that says, what's your attitude towards young and old people? You may perfectly well report, I love old people, I love young people, I love everybody! But you'll still come up with this result.

I deliberately did this one because it doesn't tend to carry an awful lot of emotional loading for people, but the reason this is a somewhat-- that's a little harsh for us old people, but the reason that this may be a disturbing test for people to take-- you should still go off and do it.

Is that if I do nice and Black African-American pictures and nasty and white, regardless of your report on-- an explicit report of what you consider your bias to be, the white population in this country will, on average, have a slower reaction time to nice Black pairings than to nice-- a pairing of nice and white.

Actually, one of the more interesting and depressing findings is that that doesn't even completely reverse with an African-American population of subjects. In the African-American population, the last time I checked on the data, that the pairings were roughly equal. So the African-American population has presumably some ethnocentric bias in its own favor and implicit bias in its own favor, but that's counterbalanced by some incorporation of the overall bias against, and so they come out as roughly equal.

So the debates in the literature about this line of work is, is this talking about implicit attitudes, which is a notion that, regardless of what we think, we're all racists, or something like that? Or is this just saying that we have somehow incorporated that we know at some level the biases of the culture as a whole even though we ourselves may not be biased?

It is an interesting question beyond the scope of what I can talk about today of whether there's a serious difference between those two. But the disturbing finding is-- and you can do this with-- it's not Black, white, old, young. This is by no means the limit on this game once you got the methodology, you can do it on anything.

So, shortly after September 11, they started doing these tests on opinions about nice-- and you don't need to do it with faces either. So if you want to do Arab versus non-Arab, you can just do it with names. So you do Abdul and Muhammad and so on, and then you do Chris and Jane and stuff, and you find that in the American population at the moment, nice and Muhammad is slower than nice and Robert.

It's not surprising, but it is nevertheless-- it's disturbing how easy it is to show these effects. And they're robust, they show up across populations, and they show up, as I say, fairly independent of what people report. It doesn't matter if you explicitly--

And let's assume that people are reporting honestly. It doesn't matter if you explicitly have the bias. You can show something that looks like a bias with a test of this sort.

Anyway, give it a try. It's interesting and disturbing. In the interesting and disturbing department, I will continue in a minute or two talking about the link between attitudes and actual behavior. So this is a good place for a short break.

[SIDE CONVERSATION]

STUDENT: With the two groups choosing \$2 or \$1 versus \$3 and \$4, my intuition tells me that even if it wasn't true, if it was just a person A and person B, you would probably get the same results.

JEREMY WOLFE: Hmm. Maybe. I'm not sure. I don't know that it's been done. My intuition disagrees, but I'm thinking-- and, I mean, I can't cite you an obvious chapter and verse, but I'm thinking that if I sat down with you and I don't take an immediate dislike to you for some reason, all right, I get \$3, you get \$4, I can feel good for both of us in some fashion in a way--

The group thing makes-- I mean, the logic-- my counter-intuition comes from all these various efforts to say, let's get Black and white kids-- or let's get Arab and Israeli kids together at summer camp because there's nothing like-- the theory tells us, nothing like getting people together to make these group identity things look a little lame.

STUDENT: I see.

JEREMY WOLFE: Now I can-- but, of course, a lot of that is, "Some of my best friends are." But for some of your best friends to be whatever, you actually have to hang out with them in some fashion.

STUDENT: [INAUDIBLE]

JEREMY WOLFE: Yeah.

OK, so I'm unusually covered in chalk dust today. I don't know what I-- oh, that's because I've been drawing-- shading things in, that's my problem.

So look, bias, we can presumably agree that bias isn't a good thing, but if it stays in the realm of private opinion, or if it stays in the realm of implicit opinion that you don't even you have yourself, it's not exactly front page news. But we also all know from reading the front page that there are regrettably frequent occasions where one group is willing to slaughter another group based on very little more, if anything more, than group identity.

So an absolutely critical question is, how do you go from-- how can people be moved from attitude to action? And the disturbing aspect of what we know from experimental psych about this is that it is surprisingly easy to have your behavior controlled by outside forces.

The experimental work, of course, cannot get people to go out and slaughter each other. Nothing like that would be even faintly moral. But rather like the Klee and Kandinsky experiments, you can do experiments that show that it's surprisingly easy to manipulate the situation in ways that changes behavior in ways that look at least a little bit disturbing.

One of the classics back in the '50s that gets this literature going was done by Asch at Columbia at the time. I think that-- I think the picture is still in the book. Marvelous collection of male Columbia nerds from the '50s. Has anybody read the chapter yet and happened to know if there's a-- all right. It's there? Somebody knows.

Anyway, look, here's the basic experiment. Here, you come into this experiment and you think you're doing an experiment on visual perception. And Asch shows you a card with three lines on it. And your job is to say which line is longer.

Now the odd thing about this-- well, you're not an experimentalist, so why should you care? But it's a little odd as an experiment to be doing this in a group. But it turns out, you're doing this in a group. And so Asch holds up the card, and you say B, you say B, and you say B, and you say B, and you say B, and you say B. Everybody says B. Fine. Next card. I'm not going to bother changing my cards, but you get the basic idea.

On the critical trial, up comes this card. He says C. He says C. He says C. She says C. C. C. Now we're up to her. I picked-- I stuck with her because she's got glasses, because the nerdy Columbia guy has glasses on, too. What does she do?

Well, why did all these guys say C like some kind of morons? They're all confederates of the experimenter. The only real subject is this person. And the question is, does she say C? The answer is, about a third of the time in the original Asch experiment, the answer is yes, she says C.

It is completely clear that even when she doesn't say C, she's uncomfortable. This is an experiment on peer pressure. And it's perfectly clear that when everybody else is saying C, she is busy taking off her glasses and checking them and stuff like that to see what's going on here, there's something wrong.

What do you think manipulates the-- so the standard result is you get about a third of the people complying with the pressure. What reduces that compliance, do you think? Yep?

STUDENT: How close they are to the same length?

JEREMY WOLFE: Yeah-- oh, sure. If we make-- then presumably it's hard to get the result in that. But-- OK. But if we keep the physical stimuli the same, yes?

STUDENT: If someone picks A.

JEREMY WOLFE: If somebody picks--

STUDENT: --somebody else picks A--

JEREMY WOLFE: Oh, if somebody picks A. Oh. And it just looks noisy. Actually, I don't know if they ever did that particular manipulation. That's an interesting question. That might change things. Yeah?

STUDENT: If they have like six or seven people who said C, so one other person said C?

JEREMY WOLFE: Yeah. It doesn't take any support. You know this-- you probably know this from arguments with groups of people or something. It's hard to be the first person to voice the minority view. It's much easier to be the second person. I think I actually have the data from that. Yeah. One supporter. One supporter in the group drops compliers from 1/3 to 1/12. So big drop in the amount of compliance.

The more people who say C, the more likely you are to comply. The smaller the group, the less likely you are to comply. But the point is, that even in a matter as seemingly straightforward as, which line is longer? You can feel that pressure from others.

The most famous experiment in this canon is an experiment done by Stanley Milgram. And in that experiment-- here's the setup. You come into the experiment for a-- into the lab for a study on learning, you're told. The effects of punishment on learning.

And there are two of you. And one of you is going to be the learner and one of you is going to be the teacher. And we're going to decide this randomly. Flip a coin. Oh, OK. You're going to be the teacher today.

Now, in fact, this is not random. The subject is always the teacher, the learner is always a stooge of the experimenter. And here's what happens. You're told you're going to do a task, and your job as the teacher is to give the learner a shock every time they make a mistake.

This is done back in the '60s with this great, big hunk of electrical equipment with a gazillion switches on there running from 15 volts to, I think, 450 volts in 15-volt increments with instructive little labels like Mild, and then up here somewhere is Severe. And by the time you get up here, it's labeled something like XXX.

[LAUGHTER]

And the instruction is that-- the rule is, every time you make a mistake, you increase the voltage one. Now-- and we will give you a 45-volt shock. You the teacher gets a 45-volt shock just to see what it's like. And a 45-volt shock from this apparatus is mildly unpleasant. It's nothing you'd want to sign up for. It's not going to kill you, though the suggestion is that this might. And that's the rule of the game.

Now, before doing the experiment, Milgram went and asked everybody under the sun what the result would be. And where's the answer to what people said? Well, the answer is that everybody under the sun said they'll bail out pretty early here, that nobody's going to give them very massive shocks. Didn't I write down a number for-- [MUMBLING]. No, I didn't.

But he asked theologians, he asked psychologists, he asked people off the street, and everybody agreed that this is not going to lead to much in the way of shocks. And this made the result of the first experiment a little surprising. Everybody went all the way through to 450 volts. The entire population of subjects in the first study went through to 450 volts.

Now, were they thrilled about this? [GLEEFUL VOCALIZING] Oh boy! No! It was also absolutely clear that the subjects were very uncomfortable about this and that they questioned whether they should do this.

And Milgram had an absolutely stereotypical-- I mean, he'd prepared in advance the response. Please continue, the experiment must go on. And that was it. You were free to leave, though he didn't explain this to you in great detail. But if you said, "Should I do this?" He said, "Please continue, the experiment must go on," and people did.

Now, he was a little surprised. In the original version, the subject-- the alleged learner had been taken get out and put in a different room. And Milgram figured, well, look, maybe these people just didn't believe the setup story here. So they moved the learner-- the alleged learner to a position where he was visible and making noises about this. And so vigorously protesting as the voltage gets larger.

At some point, up here, he says he's not going to respond anymore. And you're not going to respond anymore-- so the experiment's rigged, so that he keeps making mistakes. As a result, no response is considered an error, and Milgram is there saying, "Please continue, the experiment must go on." Oh, he also makes useful comments like, "I have a heart condition" and stuff like that. It's pretty vivid stuff.

So what happens in that case? Well great, no longer do 100% of the subjects go all the way through to the end. Only two-thirds of them go all the way through. With a subject who has stopped responding and has nouns-- for all you know, you're killing this guy, perhaps. This is pretty disturbing kind of stuff. It was very disturbing to the subjects.

This is-- actually, this is one of the experiments that produces the need for informed consent in experimental psychology. I mean, you didn't just go and hijack people off the street and say, "You want to be in my-- you have to be in my experiment." These people were volunteers, but they weren't-- there was no consent process the way we would have today. And the level of distress produced in these subjects was part of what drove the field to require informed consent.

Why was the experiment done at all? This is an experiment done in the '60s less than a generation after World War II, and a question that had obsessed social psychology since World War II was, how could the Nazi atrocities have happened? Who were the people who went and killed millions of other people-- not the soldiers, but the people who killed 6 million Jews and-- I don't remember, a million and a half gypsies and some large number of gays and so on? Who were these people?

There was a theory out there that encapsulated in a book called *The Authoritarian Personality*, which, among other things, fed into stereotypes about the Germans, which was that there was a certain type of person who was willing-- the Nuremberg trials, the war crime trials after the war had produced over and over again the line, "I was just following orders."

And the notion was, well, there are some people who are just good at that, they just follow orders, that's-- and the rest of us were not like that.

Milgram suspected that that was not the case. Milgram suspected that the answer was that under the right situation, that many people could be pushed into acts that they would objectively think were impermissible. This is his effort to get at that question.

And-- well, if this sounds like current events to you, every social psychologist with half a credential was on the news after the Abu Ghraib prison scandal earlier in the year because it just sounded so much like this problem again. The people who were charged, many of them responded with the response that they were simply, if not following orders, following the implicit instructions that they felt around them.

You got endless articles in the papers saying, oh, so-and-so was just like everybody else back home. I don't understand how he-- I don't understand how she could end up in these pictures doing things that any reasonable person would say are unacceptable in a military prison situation. It's the same kind of question, different scale of magnitude, of course, to the Nazi atrocities, but it's the same question-- how do people end up doing things like this?

Before attempting to answer, which I can see, is going to run into my next lecture, let me tell you about a different experiment designed to get at the same question. This is an experiment where you're in one of these-- you think you're in a study-- a consumer relations kind of study that you might run into at the shopping mall. I mean, they set up the experiment at a shopping mall.

And the cover story is this. We're doing some research on community values because we're basically doing investigations for a legal case. Here's the situation. This guy was living with a woman to whom he's not married. His employer found out about it and fired him. This has gone to court, and the issue is whether or not-- the court case hinges on community standards.

If community standards are that living in sin out of wedlock is a bad, bad thing, then it's OK to fire him; otherwise not. So we gotta find out what the community standards are. Let's all have a discussion.

So, OK, I've told you the story. I've got the cameras rolling here. I'm going to-- I, the experimenter, I'm going to step out and you guys discuss. OK, you guys discuss, that's great.

Now, I come back in. There's a group of, say, 10 of you or something discussing this. I come back in and I say, that was great, thank you very much. Now, I know what you believe because I've been watching this, but I want you, you, and you, please to argue from the point of view that the guy should be fired. I know it's not what you really believe, but just argue for that.

All right, fine-- [MUMBLING]. Comes back in. OK, now I want you, you, and you to join that argument arguing that the guy should be fired. Fine. And then in the final-- the penultimate step is, I want everybody to have a chance to look into the camera and say why the guy should be fired. I know you don't believe that. Whatever. But just give me a little speech as if you believed it, OK? Cool.

OK. Last step. Here's this statement that says that I can use my videotape in any fashion I wish, including submitting it as evidence in court. Will you please sign? Ah. I'll be back in a minute. I gotta go rinse a few things out, but you guys decide, is the way the thing works. The question is, do people sign?

I mean, this is obviously-- basically asking you to perjure yourself if that wasn't what you believed. Now this was a huge experimental design originally. They were crossing a number of people with-- they had gender and lots of-- they originally had a plan for-- oh, I can't remember, a gazillion groups.

They called off the experiment early because like the Asch experiment and like the Milgram experiment, this experiment produced extremely strong feelings in their subjects, and enough that they had gotten a form of informed consent, which I can't go into now, but even with that, they realized that it felt unethical to them to keep going.

So they had 33 groups. It was a busted design, but they had 33 groups. Of those 33 groups, how many did the Milgram thing, went all the way and everybody signed, do you think? The answer is one. Only one, I think. One group got total obedience. 16 of the groups got unanimous refusal to sign. Nine groups got majority refusal to sign.

So in this experiment, compliance was much, much lower. And the question that I'll take up next time for the start of the next lecture is, what's the difference between the Milgram experiment-- the Milgram experiment-- let me say one last thing about the Milgram experiment.

This is not an isolated result that happened-- Milgram's at Yale. It's not just an isolated experiment that happens in New Haven in 1960. Replicated all over the place. Doesn't matter in America, it doesn't matter age, sex of the subjects. It replicates beautifully. Why does this work and the other experiment didn't?