# Right vs. wrong: The psychology of trust

**Host:**          **David Mallon**, vice president and chief analyst, Bersin, Deloitte Consulting LLP

**Guests:**       **Jim Guszcza**, US chief data scientist, Deloitte Consulting LLP
                    **David Pizarro**, associate professor of psychology at Cornell University;
                    chief science officer, BE Works

**David Mallon:** Welcome back to the Capital H podcast, where we explore the topics and trends related to work, the workforce, and the workplace. I'm your host, David Mallon.

Today, we're bringing you a special episode. Each year, Deloitte is part of an annual behavioral and data science symposium called Nudgeapalooza. And today we're excited to share part of that event.

We all know that ethics, morality, and trust play a huge part in how we do business every day. In life, in work, we make assessments about whether people we encounter seem trustworthy or moral. Moreover, consider the scrutiny that organizational leaders face for certain decisions they make—the way they communicate those decisions to internal and external stakeholders seems, at times, as important as the decision itself. Researchers such as David Pizarro, our guest on today's

episode, have done extensive research on the psychological factors that underpin these assessments of trustworthiness and morality in others—particularly when the decision is difficult or there's a lot at stake.

Before we jump in, I'd like to welcome my colleague, Jim Guszcza, the US chief data scientist for Deloitte Consulting. He's going to give us some background on Nudgeapalooza. Welcome, Jim.

**Jim Guszcza:** Thanks, David, glad to be here. So the Nudgeapalooza is an annual behavioral and data science symposium hosted together by Deloitte and Georgetown University's McDonough School of Business. And the idea is, each year, we bring together academics, business leaders, and policy experts to share and learn about the latest insights and strategies for ethically "nudging" positive behavior change.

The theme for Nudgeapalooza 2019 was "Ethics in a Data-Driven World," and it featured discussions on the intersection of behavioral insights and artificial intelligence, exploring the nuanced ethical considerations that leaders encounter as capabilities in both of these fields—behavioral insights and AI—advance. And by way of background, "nudges" refers to designing choice environments in ways that remove barriers and go with the grain of human psychology. The goal of choice architecture or nudges is to prompt better behaviors and smarter decisions on the parts of citizens, employees, patients, customers, and so on. And as nudging becomes more common, both in and out of government, so do questions about how behavioral insights can be used both ethically and unethically. That was a big theme at this year's event.

**David:** Thanks for that background, Jim. When will the next Nudgeapalooza be, and how could our listeners learn more?

**Jim:** So it's still unofficial and, as the name suggests, the Nudgeapalooza is a little informal, but we think it'll happen again, I'm guessing in December 2020, probably again at the Georgetown McDonough School of business, and listeners can learn more by listening to the Capital H podcast.

**David:** Great. Thanks so much for joining us today, Jim. Like I mentioned, in today's episode, you are about to hear a talk from David Pizarro, recorded at the most recent Nudgeapalooza. David's an associate professor of psychology at Cornell University and chief science officer at BE Works. I hope you enjoy.

**Jim:** I am delighted to announce our next keynote speaker, and thanks to Jackie Stein

for making the connection. This is Professor David Pizarro from Cornell. David is a professor of psychology at Cornell. He is also the chief science officer at BE Works. I have known about David's work for a long time, I have listened to his talks for many pleasurable hours on edge.org, so you can check those out—do it. David also has a podcast. He is one of the top behavioral scientists working in the area of moral judgments and moral intuitions and the biases that effect moral judgments. So please welcome David, who's going to speak on the psychology of trust, character, and moral attribution. Thank you.

**David Pizarro:** Thank you so much. It's my pleasure to be here. There is a fundamental question that we have to ask ourselves whenever we interact with somebody, especially with somebody whom we don't know, for the first time. We want to know, "Can I trust this person, who can I trust, who is a good person?" whether it's engaging a novel business project, or dating somebody, or just meeting somebody and deciding whether or not . . . if they're safe to sit next to one on a bus. We are motivated to distinguish good people from bad people, but the truth is, humans have a unique feature—we are different than every other animal on this planet in many ways, but in one specific way— it's not just our intelligence that has gotten us this far; it's our desire, our willingness, and our ability to cooperate with each other. So scientists have called the human species hyper-cooperative. We really like engaging in cooperative activities with each other. Because of that, we can do things like build a space station, put together a conference. If it weren't for our ability to trust that others are going to fairly and willingly and helpfully engage in a cooperative task with us, we simply wouldn't be able to do this.

The reason that cheating is an interesting thing is that it tends to be an exception to all of the behaviors that we generate. So if I asked you right now, would you take $2,000, just free, if I gave it to you, you would probably say yes. Rational economic model, $2,000 versus zero is the case. I have left my laptop on my chair, along with my iPhone and a bunch of other valuables in my backpack, I am not worried, but anybody could easily

walk in here, take that, and I would not notice, but we don't do that. We are hyper-cooperative. What does this mean? It means that we get to do things that again no other species gets to do, but as evolutionary biologists and game theorists have pointed out, this leaves us very vulnerable, because we go around trusting people to do things that will help us and that won't harm us.

So, for instance, Amazon delivers about . . . At least a couple of years ago, it delivered 5 billion packages in one year. Now you could just go to any neighborhood and walk around and pick up packages, but most people don't do that. If most people did that, Amazon would not work, but some people do, and because we are so trusting and so cooperative, we assume that another person won't do this, but in fact, they do, and it's our niceness, it's our trust in other people that allows certain individuals to manipulate the system, to game the system. So if you are willing to do that unethical thing, you can get far.

I want to talk a little bit about some experiments that we have done. Before I get there, I want to tell you there is a deep and rich tradition in psychology that tried to look at this question of "Who can I trust? How do I know that this person is a good person?" This is from Vaught's Practical Character Reader from 1902. I don't have the text here, but he says that he had millions and millions of observations, and what he would do is describe the shape of people's bodies, their heads, their facial features, and induce from those that people with certain kinds of physical pictures would be honest. This is the science of physiognomy. It has an old and rich tradition.

More recently, people have actually shown that we do use facial features in order to determine who is trustworthy, but they are not accurate. How close somebody's eyes are together are not going to really tell you whether or not they are trustworthy. We are just biased in this regard, but there are some clues to answering the question of "Who can I trust?" I want to tell you, today, the behavioral science has come some way, it is a difficult question to answer, but we have come some way in giving us clues about who we can trust

or at least telling us how people who are trustworthy act and how we get judged by others to determine whether or not we are trustworthy or good. So first, I want to talk a little bit about nonverbal cues. There is a rich literature on this as well. I am going to tell you about a study that we did. This was with the economist Robert Frank, psychologist David DeSteno, and roboticist Cynthia Brazil from MIT. We did a couple of studies. So what we did first is bring people into the lab, and they were either face-to-face with each other on little tables, they sat, and we had cameras, three cameras capturing three angles of their interaction. We allowed them to interact face-to-face for five minutes, or we brought people into the lab and had them chat with each other over the Web. In both cases, it was two strangers talking to each other. We told them that they were going to play an economic game. This was called the give-some game. It is sort of a version of a trust game. What we said is, everybody starts with four tokens. Each of those tokens is worth one dollar. So if you turned it in at the end of the experiment, you got four dollars if you had four tokens. Here is what they could do: They could give those four tokens to their partner and they would double, leaving their partner with eight dollars, and if you engaged cooperatively, if both people did this, both people would leave with eight dollars. Here is what you could do as well, though. You could convince the other person to send their four dollars, thus doubling it, but keep the four dollars for yourself that you started with, leaving you with 12 dollars. So you are, in essence, deceiving the person, trying to tell them that "You are going to trust me," and then keeping the money and the 12 dollars. So the selfish option gets you more than the cooperative option, so long as you can tell the person . . . if you can convince the person to cooperate with you. So here is what we found when we looked at face-to-face interaction versus Web chat. People in both conditions gave the same amount of tokens. That is, they were trusting in the same level, but we asked people to predict whether or not they thought their partner was going to return/reciprocate, and here is what we found. This is plotting prediction errors. When people were face-to-face, they were more accurate in their prediction

as to whether or not the other person was trustworthy. So they were getting some information face-to-face that they were not getting over a Web chat. What information was that? We had videotapes of all of their interactions, so we put undergrads to work, hard-coding all of these. This is a very labor-intensive task to code every single gesture that people used, and in a very bottom-up way, in the statistical way, we found four gestures that predicted whether or not somebody was going to cooperate. These were hand touching and hand rubbing, kind of like the old-timey villain; face touching; crossing of the arms; and leaning away. Now, I don't want to overinterpret it, but they are all gestures that in some way put some distance or something in between you and the other person. We have no idea whether people consciously knew that they were reading these nonverbal cues. We just knew that these were predictive of cooperation. So these seemed to be the gestures that were differentiating somebody who gave versus somebody who didn't. So we wanted to use these gestures and conduct a true experiment, because these were studies where both people were naturally in the study and were generating these nonverbals spontaneously.

We wanted to do an experiment where we could manipulate whether or not somebody used these gestures or not in a clean way. So we went to a robot. This robot is Nexy. Nexy is a social robot that can make many facial expressions and bodily movements. We programed Nexy to interact with participants. We brought participants in from the Boston area. We told them they were going to play that exact same game, and here is what we did. Because we can program this robot to do whatever, for half of the people, we made Nexy do those gestures, those gestures that we found in the first experiment, and for the other half, Nexy didn't. I'll say "she" because the voice of Nexy was a woman in this experiment. We had actually two graduate students, one who is running the macros for the movement, the other one who is talking. Participants were actually weirded out a bit, but within minutes, they were talking to Nexy as if she were a person. In fact, some people walked

out and they said, "Well, I can't believe that robots have gotten this smart." It was a grad student, that's why.

When we asked them to judge "how trustworthy do you think this robot is?" after their interaction, people in the control condition judged her to be more trustworthy than in the condition where she was generating the nonverbal cues that we had programmed into her, and more importantly, they actually played the game with Nexy, and in the control condition, they gave more tokens than in the negative condition. So they were more trusting both by self-report and behaviorally. This is just one hint that nonverbal gestures do, in some cases, communicate whether or not you can trust somebody, and chances are, we are reading them. We are at some level reading the interactions for nonverbal cues, and we are not bad. We showed at least some success. So let's move from nonverbal cues to something probably a bit more obvious: the expression of moral intuitions. What do you find morally intuitive? That says a lot about you. If you spontaneously express that you find something to be very wrong, then I might think of you as a good person. If you have weird intuitions, or if you don't share my intuitions, or if it takes you a little bit of work to even get to that intuition, then we are a little bit more suspicious. So let me give you an example. You may have heard of this; this has been talked about quite a bit. This is an example of what we call a sacrificial moral dilemma; this is the trolley problem. A runaway trolley is heading down the tracks for five workers, who will all be killed if the trolley proceeds on its present course. Adam is on a footbridge over the tracks in between the approaching trolley and the five workers. Next to him on the footbridge is a stranger, who happens to be very large. The only way to save the lives of the five workers is to push this stranger off the bridge and on to the tracks below, where his large body will stop the trolley. The stranger will die if Adam does this, but the five workers will be saved. It is a classic sort of moral dilemma, ethical dilemma from the literature.

One time I was giving a talk—Cornell has a prison program, where they get graduate students to teach courses, give credit to prison. It is a maximum-security prison in Auburn, and they asked me to give a lecture. I was very nervous; for one, I didn't realize I couldn't bring my laptop and my slides, so I had to just impromptu. This was about 50 inmates who are in there, in many cases, for very, very serious crimes, and some for life. I was talking about this, I gave two examples, one where you have to flip the switch, which most people are okay with, and this one where you have to push someone to their death. So I asked the inmates, "Do you think it is okay to push the person to their death?" and most of them said no. They all had that intuition, except for one guy. And he goes, "It's the same thing, but you know, in one case you push the guy and in the other one you just flip the switch. It's the same thing." And at this point, the people around him are getting nervous, because these inmates are actually in this program because they are so well-behaved, but if they do any bad thing, they will get kicked out of this program. I think the other inmates were a little bit nervous, but the guy, in some sense —many philosophers would agree—was right. What's the difference between how that person dies if you are sacrificing one to save five? Well, anyway, most people's intuition is that in these kinds of cases—and we gave 14 different cases, all of which would require the sacrifice of an innocent person to save more lives—what characterizes somebody who says it's okay? Well, we know a bit, right? The people who kill or are willing to kill one to save five, this is called the consequentialist or utilitarian. The people who are unwilling to kill an innocent person to save five, we call them deontologist. These are people who believe— there are some constraints—"It doesn't matter how good the consequence is, I cannot violate the rule of killing an innocent person." So what characterizes people who are were willing to say that they will kill one person to save five? Well, in one study that I did, we found that people who are willing to say that they will kill one to save five score higher on measures of psychopathy and measures of Machiavellianism. These are people who are willing to adopt the counterintuitive,

right? Consequentialism is a very difficult ethic to really endorse, to really get behind, because it requires you to do some very uncomfortable things or at least agree that they should be done. But what we found is that people who readily endorse it are higher in psychopathy and Machiavellianism. This is consistent with a whole bunch of other literature showing that it's our emotional response that is preventing us from saying we will push somebody off the bridge. It's that feeling of restraint at the thought of pushing somebody that is generating the intuition that it's wrong, and that's a good intuition to have, right? We don't want somebody who might push us off the bridge when we are walking because they see five people. You want your friend to be loyal, you want your friend to not kill you as soon as he calculates the outcome, which I will get to when we are talking about computers and AI. In fact, in other work from Fiery Cushman at Harvard University and his colleagues, they show that responses to these kinds of dilemmas actually predict a participant's willingness to engage in fake but very realistic violence. They get a real-life baby doll—like probably one of those used in CPR, looks really, really realistic—and they ask the participant to just grab the doll and smash it on the table. Now, you probably had an emotional reaction to that; people do, even though they know it's a doll, it is a little bit hard to do when it looks so much like a baby. They had a bunch of these tasks, including . . . The experimenter himself sat down and put up a fake plastic leg that looked real and said, "Just take this hammer and hit me as hard as you can."

They hooked people up to physiological measures, and what they found was that people who were willing to say that they would sacrifice an innocent life had less of an emotional reaction when they were engaging in that violence; moreover; they actually hit harder than deontologists. The deontologists had this restraint. So what your moral intuition is . . . If you generate that quick moral intuition, like the guy at the Auburn State Prison, that makes you a little suspicious. In fact, in work that I have done with Molly Crockett and Jim Everett, we had people engaged in a trust game with each

other, not unlike the game that I showed you, but before they did, we told them what somebody responded to. We said, "Either this person responded, 'It's better to save five lives, so of course I would push the guy off the bridge,' or they said, 'No, I can't do that, that's just wrong.'" So now, pretend you are a participant, you are about to engage in a cooperative trust game. You have heard that somebody thinks that it is okay to push one person to save five or you have heard them say, "No, it's wrong." Well, it turns out we trust consequentialists less. Deontologists, people who say that it's wrong, are perceived as more moral, and they are perceived as more trustworthy. When we actually have them play the game, people give more money to the deontologists, and they predicted they will get more money back from the deontologists. That is, there is something about the counterintuitive nature of expressing a consequentialist judgment that makes people suspicious of you. So Robert Frank, the economist, was once telling me that he was a utilitarian at dinner—and he was at Cornell—and I said, "Well, you can't really be a utilitarian. You wouldn't push somebody to their death." He says, "No, I wouldn't. I think it is right; I just wouldn't want that person on my team." And that is the assessment that we are making. Somebody might be a good person because they can adhere to their utilitarian principle; I just don't want them around. If they are calculating, they might calculate me out of the equation. We want our friends to have a particular set of emotional reactions, of moral intuition, that tell us that they are a good, trustworthy person that we can engage in whatever activity.

Finally, I am going to tell you about some research showing that it is not just what intuitions or judgments that you express or make, it's how they are done that often matters. In fact, how you make a decision can often overwhelm the "what" of your decision. So we gave everybody a scenario describing a hospital administrator. This hospital administrator is faced with an ethical dilemma. They must decide whether to fund or deny a costly liver transplant for Johnny, who is a dying five-year-old patient,

or they can use that money to buy more equipment and hire more doctors and save more lives in the future. So what would the hospital administrator do, what ought they do, what would you do in that situation? So here is what we told participants. Some participants were told that the administrator decided to save the child, or the other group were told that the administrator decided to buy the equipment. Buying the equipment is a more consequentialist response; buying the equipment is seen by most people as not a nice thing to do. Most people say that the hospital administrator should use the money to save the child. But here is the other thing we manipulated. We manipulated how that administrator made the decision. So we told half of the people that he deliberated carefully before arriving at the decision, or that he arrived at the decision quickly and easily. Now imagine what we are saying: Just like the guy in prison, if you immediately express the consequentialist intuition, that says something about you. So everybody likes it when the administrator decides to save the child—actually, it doesn't matter whether you make that decision slowly or quickly—but if you decide to buy the equipment, that counterintuitive judgment, the more difficult judgment, if you make that decision quickly, people don't like you as much as if you make that that decision slowly and carefully. In fact, there is no real difference between the first three bars, saving the child, buying the equipment, slowly and deliberately. So in this case, you are communicating that the decision was tough, and that you thought carefully about it, and you then made the decision, because if you make that decision quickly, it says something about what kind of person you are. Where do those judgments come from? Do they come from a place of the natural sentiments that Adam Smith talked about, the moral sentiments? The tender other feeling human emotion, or those active in you? Or are you a cold, calculating machine who's willing to do whatever?

Think about this: I have a friend who was diagnosed with cancer, and he went to the doctor, and the doctor said, "You know, you have thyroid cancer; great news, though: That is the best kind of cancer," and he

was so angry. He said, "How is she going to tell me that it is a good kind of cancer?" He is finding out that he got cancer for the very first time, he is devastated, and she is nonchalant about it. The doctor is in an awkward position. They are facing their patients all the time, and she sees people she has to give devastating news to every day, so this is good news for her, but if I am going to give any advice, it would be, when you are a policy maker or a doctor or anybody in a leadership position who has to communicate that a tough ethical decision was made, I think it is good for you to express those emotions, even if they are not sincere at that point. That is at least what I think the literature is telling us. This is why I believe that some people have problems with artificial intelligence making ethical decisions. I don't think that anybody believes that a computer is computationally incapable of taking the information into account and generating what a human would generate; it is an ethical decision. Think about it; it is not that hard, where we are oftentimes using intuitions, we could program that into a computer. What I think is wrong, where I think people think is wrong with an AI making an ethical decision, is that the computer doesn't arrive at their moral judgment in the same way that you and I do. We have those emotions that tell us "This is the right thing to do; this is something I should never do." Programming that into an artificial intelligence is not giving us the same information that you would give me if you were expressing your moral judgment. It is a robotic way, a hyper-rational, cold way of arriving at a decision. Even if the decision matches, the "how" of the decision makes a difference. This actually has implications for the effective altruism movement, and I have talked to a few people who work at effective altruist institutions. They are faced with a problem. They believe that the best way to determine how to donate to charity is to run the numbers, and when you run the numbers, you realize that some charities can save way more lives with the same amount of money as other charities. People just don't like that approach. We want our moral judgements to be made with a certain amount of emotion behind it.

So I want to end with just saying—like from the literature—suppose that you do want to communicate your good character. Here are just a few things. In person, we can give a lot more information face-to-face than we can over email, over text. Express your moral intuitions and your commitment to the intuitions that everybody else has, the shared rules and values; demonstrate that you have those tender human emotions, compassion, and empathy; but if you have to make an ethical decision that is difficult, that is counterintuitive, that goes against what some people's intuitions are, express that you made it with weeping and gnashing of teeth, that it was a tough decision to make, and even if it's not and if all fails, fake that sincerity, find a way to communicate over and over again that you have these emotions, that you went through this deliberation, even though you have done it a million times. With that, I will end. Thank you for your time.

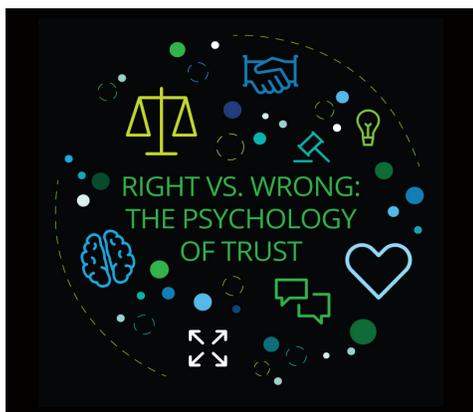**David Mallon:** What a fascinating talk, Jim. What stands out for you most?

**Jim:** Well, we felt very fortunate to have David Pizarro join us. I've been following him for years, and I consider him one of the foremost psychologists working in the area of psychology, morality, and ethical judgements. David's theme of trustworthiness is so incredibly timely, both from the perspectives of future of work, but also artificial intelligence. Now, I think David said something earlier in this talk that really . . . When people accomplish things, whether it's making a sandwich or building a building or completing a project at a consulting firm or completing a government project, it's never one person. It's a team of people working together, and trust is kind of the glue that holds teams together. It's like bonds that enable molecules and things to come together to create smart teams, and so without trust, that kind of group intelligence goes away. So that's from a future of work point of view. We need to create work environments, which are increasingly becoming digital, and in which people can trust each other and work together.

It also relates a little bit to AI; a lot of people are talking about trustworthy technology and trustworthy AI. And really, the right way to think about AI is not building smart machines, but building machines to make human groups smarter. So we really need technology we can trust. We need trustworthiness in the part of people and technology. So it's very fortunate to hear David's perspectives, and I think it's one of the major themes of the day.

**David:** Trust, particularly in the relationship between humans and machines, is fascinating, and it's of course very important, and it's so interesting to learn about the psychology that underpins our perceptions, our decisions—and the research methods that David and his colleagues have used to uncover these kinds of insights. I hope you all enjoyed this as well.

Thank you again to David Pizarro and Jim Guszcza for joining us on today's episode, a special feature from Nudgeapalooza '19.

Please join us next time on the Capital H podcast as we dive into more topics and trends that focus on putting humans at the center of work.



RIGHT VS. WRONG:
THE PSYCHOLOGY
OF TRUST

## Learn more

Visit the Capital H podcast library
[deloitte.com/us/capital-h-podcast](deloitte.com/us/capital-h-podcast)

Explore our blog for additional insights
[capitalhblog.deloitte.com](capitalhblog.deloitte.com)

Join the conversation on Twitter
[@DeloitteHC](@DeloitteHC)

This podcast is produced by Deloitte. The views and opinions expressed by podcast speakers and guests are solely their own and do not reflect the opinions of Deloitte. This podcast provides general information only and is not intended to constitute advice or services of any kind. For additional information about Deloitte, go to deloitte.com/about.