Recruiting Talent for a Changing Workforce

Host   **David Mallon**, Vice President and Chief Analyst, Deloitte Consulting LLP

Guest: **Jason Dana**, Professor at the Yale School of Management

During this last decade of economic recovery and expansion, organizations are focused on finding the right talent to drive business growth, but with record low unemployment rates and skill shortages in many technical areas, talent acquisition has only gotten harder. Compounding the problem, most of our current recruiting processes are actually holding us back, preventing organizations from finding the right candidates. Our guest today has researched decision-making in a wide variety of managerial contexts, including hiring. He is Jason Dana. He will talk to us about how organizations can rewire their recruiting model and their recruiting mindset.

David: Hello and welcome to today's episode of our CapitalH podcast. Today our guest is Jason Dana, who is a professor at the Yale School of Management. Jason, welcome to our podcast.

Jason Dana (Jason): Hi, thanks for having me.

David: Thank you. We are here to talk about, I guess your research in general, but let's start with some research you described in a recent New York Times Gray Matter column about, I think the title of your column was called, "The Utter Uselessness of Job Interviews."

Jason: Sure. Just by way of context, there is a fair amount of research in industrial psychology that shows that interviews aren't very good, especially like unstructured interviews, ones where the interviewer just makes up whatever questions they want. At least they are much less good than people think they are. So what we were trying to get a handle on basically is why interviewers when the evidence shows that they are not particularly good at predicting anything, they are not particularly effective, why do interviewers tend to like them so much and why do we use them so much, why do we believe in them so much. We identified a couple of reasons, and I will tell you about the studies we ran. We basically had some student subjects make predictions on a topic which they know about, which is predicting another student's grade point average for a given semester. So to make this prediction, they were given some background information like their course schedules and things about them and really good information, which was their prior GPA. And so one of the things we found is that their predictions were worse when in addition to this background information, they were also allowed to perform an interview of the student and ask them some questions. This actually hurt their predictive performance. What we call this, while its effect in social psychology has been identified called dilution. So, the dilution effect is the tendency for irrelevant information or noise to reduce the impact of valuable information. They started off with some really useful information in terms of course schedules, prior GPAs, and then you might get some more good stuff from doing an interview, but you also might get some noise or some nonsense and rather than simply being ignored, extraneous information tends to hurt predictions by making us make less use of the good stuff.

Now unbeknownst to our interviewers, the second thing we found is that, so some of the interviewees were actually answering questions according to a pseudo-random system that we came up with and it basically just relied on the first letters of the last two words of a question. So we divide the alphabet A–M, N–Z, and if both of the first letters come from the same category, A–M or N–Z, then they should answer yes or in the affirmative, and if the categories don't match, they should answer no. What's perhaps even more interesting, the predictions after the random interviews were no worse or better than after real interviews, but maybe even more interesting is our subjects actually rated, we asked them some questions after they did the interview and made their predictions, and they rated the extent to which they got to know the person and the extent to which they got useful predictive information from the interview, these ratings were as high or higher after random interviews. So we call that sense making. Sense making is a tendency to form a coherent narrative out of what happens regardless of whether it actually makes sense, regardless of the stimuli.

David: We call it sense making but it might really be nonsense making in this case.

Jason: Yes, nonsense making, right. So another problem then is that we make too much sense of everything and thus we have a confident feeling of understanding people even when we don't understand.

David: Let's go back to the first part of your, the first study we discussed, the dilution effect and the influence of noise on peoples' decisions. So GPA is the actual predictive factor, the strong predictive factor of future grade point average, historical grade point average, and people were convinced that by having this interview they got all this other information and it sounds like they overweighted it, right?

Jason: Right. Hard to say overweighted because we don't know what they are doing in their head, but what we can say is that look if I just handed them the background information, obviously the predictions they would make about the person would be pretty highly correlated with something like prior GPAs because they don't have much else to go on. You start giving interviews, which of course they don't have to use, they don't have to weight them heavily, but they get all these other things to keep in mind and then I think what you would see happen is that their predictions after the interview would be much less correlated with some of the background information like prior GPA. So in a sense, what we can say is you are diluting the potency of this super good predictor by giving people a bunch of other stuff that may or may not be useful.

David: I want to drill more into the human resources hiring and talent management implications of this finding, but for a minute would it be okay if we sort of generalized a little bit to decision-making in general? I have always thought that findings like this have just incredibly far-reaching importance for all aspects of business, especially when considering using data to drive better decisions. It seems like what you are uncovering, and you tell me if I am getting this right, is that humans are kind of bad at weighing information in kind of a coherent or optimal way when making case-by-case decisions. Is that a fair characterization?

Jason: Yeah, and actually there's a, I'm actually stealing your words I think now. There is a venerable tradition in psychology research comparing pattern of the forecast accuracy of human judges who may come up with things implicitly or in the head or holistically and then formulas, even simple formulas, and this research far predates what we now think of as artificial intelligence and having big data sort of stuff. So Paul Meehl wrote a book that I think we both really like back in 1950s on clinical versus actuarial prediction. So he had 20 studies that had compared in some way actuarial judgments to human beings across a broad variety of domains. In all of them, the formula either tied or beat the human judges, and this caused, I guess, apoplexy among psychologists that we can run that total into the hundreds and we can say then in the overwhelming majority, the formulaic approach tends to out-predict the clinical approach, if you want to call it, the holistic human judgment approach.

David: Right. And that could be hiring decisions, judging the quality of wine, insurance underwriting decisions, scouting baseball players. The movie Moneyball is an illustration too, right?

Jason: Who is going to win a sports game, how long somebody is going to live.

David: So let's get back to the human resources context. There are all these famous headlines that are in the news these days and there is one in particular that a lot of us have in mind, which is there is a prominent example of a company that built a machine learning base, big database, hiring algorithm, because they need to make a lot of hiring decisions and they never used it. They pulled the plug on it because apparently they audited their own algorithm and realized that it gave negative weights to words to noting the person who is applying for the job is a female. So it was biased against women, in other words. It is an interesting result on many levels. There's this increasing awareness that if you train a machine learning algorithm on data that contained biases in the decisions, you will just get an algorithm that encodes and propagates and maybe amplifies those biases going forward.

Well, an interesting implication of what you are describing is that there are other kinds of algorithms, a machine learning approach is not the only way to get an algorithm, it might be very useful in a business context. So it's one implication.

Jason: When I talk about this stuff in classes, one of the points that I hammer home a lot is that as much as anything, preparing a model of a decision disciplines your thought. It makes you think through to be explicit about what you value and what you are looking for, and you could think of different ways to do this. You could just sort of look at resumes and have a little committee and I guess you could discuss it and say, well, I like this, I like this, or you could try to make a model, but if you try to make a model, the first thing you would be confronted with is what am I trying to predict. You have to have some sort of criterion or gold standard, if

you will. Like what would be a good admin, right? Because we won't be able to predict who is a good admin or bad admin until we can define what a good one is. And it is in that very definition that all the problems come in. So there is a lot of talk about the ethics of AI, I see a lot of confusion in these discussions because it is not AI that is ethical or not ethical, the thing that is ethical or not ethical is your objective function, is the thing you are trying to predict, is that gold standard. So if you think about the different objectives you could come up with, let's say, in admitting students into a school, some things that sound pretty reasonable could get you into trouble real quick if you haven't thought them through. So supposing I was to say, well, success is placing people in good jobs and, therefore, I am going to try to do a predictive model to admit people and predict who places well. Well, that is all good and well and it sounds reasonable but then if you think about it, you would be committing yourself to replicating any biases that the job market has. So if the job market is discriminatory against anything, the problem wouldn't be the model, the problem is that you don't have a clear objective function. There is no way around that.

David: This is not hypothetical at all, obviously. In fact, we've got, some of my colleagues right now are batting around some ideas right now to help a client who is interested in figuring out how to make better promotion decisions. Suppose that the client had an old boys' club dynamic where people went to the same social clubs or came from a certain demographic group were the ones tended to get promoted. So if you got these kinds of historical biases, it will be reflected in the outcome variables you chose. Do you have any advice as to how to go about setting these things up, choosing your objective function, choosing your outcome variable?

Jason: I wish I had a simple answer for you, I wish there was a simple thing I could say but without thinking about the proper aims of the organization and the job that you are hiring for, in other words, you need a mission, it is really hard to think clearly.

The easiest thing for people to do is just to find the people who are already at the top or who have already succeeded in the organization as the goals and then try to replicate them, but as you point out, all kinds of problems could potentially arise from that because it's a commitment to doing things the way you have always done them, which may or may not be appropriate going into the future. It could produce homogeneity, maybe you don't want too much homogeneity, et cetera, but this is never an easy question. What does an ideal hire look like? What does the ideal employee look like? That's the kind of difficult thinking that goes into the decision and the model just exposes whether you have thought that through carefully or not. The model exposes whether you really even have a clear objective and whether you really know what you are doing, but unfortunately you may be plodding along not really knowing what you are doing anyway, not really having a clear objective anyway and because you have never tried to model the problem, you have never realized how serious that step is, right.

David: That's right. So this is actually quite interesting. I'm just going to try to summarize a little of what we are talking about and you tell me if I am going astray here. It seems like we have talked about two sort of modes of human cognition. One mode is this idea of like okay, I want to decide is this building unsafe, I'm a building inspector, is this employee likely to be a good employee, is this red wine going to be a high-quality red wine vintage, and what we found is that simple formulas are better than any other judgements in making those kind of decisions. You just talked about another mode of human cognition, which it seems to me algorithms are going to fall on their face on, which is going to be using this kind of judgment to kind of think through what do we mean by a good employee anyway, kind of defining what success looks like, that is really a matter of strategic thinking and judgmental thinking, not just weighing risk factors better. So there is a case where human cognition is necessary over and above automated machine learning methods, right?

Jason: Sure. I am sure there are some AI people that will yell at us for this, but I think that is exactly right. We are saying that models are better than human judges, but if we were being precise with language, we would say the outputs of the model are better than the outputs of somebody's holistic judgment, but in all of this old research comparing models and experts, it was people that built the models, it was experts that knew what the things to include in the model were, it's experts that even decided what the outcome your prediction should be. Okay, there are now machine learning and artificial intelligence that can actually discover relationships by themselves and teach themselves, things that we didn't know about, but I think in all cases, I think I am willing to commit to this statement, it is usually the people that are telling it what the proper objectives are, governing how to get there. That's the step you can't replace. That's what people are there to do.

David: That's exactly, right and I don't think that is going to change anytime soon. By the way, Jason, as a data scientist when you said I am afraid I can't give you a simple answer, I will pay you later for that. I was applauding because I think there is this irreducible aspect of judgment. Let's keep drilling this for the human resources setting. So for an organization that wants to overcome these problems with human judgment in making hiring decisions or promotion decisions, I guess the way to approach it is not with sort of a naïve machine learning approach. Machine learning might play a role, but the answer should not be just be saying we will gather the data, we will fit an algorithm, then we will implement the algorithm and follow the algorithm, that I wouldn't say necessarily a recipe for problems, but there are a lot of red flags there with the kind of naïve approach to that. It really should be more of a judgment-driven process of figuring out what kinds of risk factors we should weigh together. And sort of like guarding four different kinds of biases. Well, what I have in mind is both an old example that I read about first in Daniel Kahneman. You know, Kahneman, he said he was inspired by our friend Paul Meehl to do this, when

he was in the Israeli army, he built a scoring model to help people decide who should be promoted from enlisted person to officer, and I think the procedure was along the lines of, based on our prior knowledge, let's talk about what success looks like for an officer and let's figure out on this dimension how do they stack up on a scale of say one to five. And maybe we have five or six or however many of these dimensions, and then if you are evaluating 20 people for five positions, try to pre-commit to adding up the scores and try to adhere to the scores as much as possible and only after you have done that, let your judgment rein a little bit more free and say, well, this person gets the highest score but that other person, the algorithm didn't know this exceptional circumstance here. So maybe we should decide and favor this person is better than that for the promotion. That's from memory, but I think that's kind of the approach he took constructing this algorithm.

Jason: I know one thing that Kahneman did was to take their procedures and instructions, which didn't make it fantastic, but it made it a little better, a little more predictive of future performance. There are a number of reasons to think that structure would be important. In the article you referenced, I told a true anecdote, a funny anecdote, a friend I had many years ago. She had an interview. She was told a time to show up, let's suppose it was 10:30, and she is in the waiting area at 10:25 and some person, the reception person, whisks her into the interview room and there's three people there waiting and they do the interview and it goes pretty well and so they offered her the job on the spot and then in the postmortem they said, well, you know the way that you could come in here 25 minutes late and still be that calm and collected, I just don't know how you do it. Of course her heart almost stops. As it turns out, she was told the wrong time by half an hour and she didn't know she was 25 minutes late. I mean, there is so much wrong with that story where to begin. First of all, you could have interpreted that exact same interview as being very negative, right, because you could say, well, she doesn't care about being late, but the other thing

that's interesting is just that okay this is an absolutely unique situation and you are kind of impressed by how the candidate responded to that particular event, but you really have no idea how other people respond, and the thing is that when you conduct the interview differently from person to person or you do whatever screening methods different from person to person, then you are always sort of doing an apples to oranges comparison about people. You can't compare the quality of someone's responses to everyone else on the exact same prompt because they didn't get the same prompts. So this is generally not a good thing, for fairness reasons but also just for accuracy reasons, it's often best to compare apples to apples. So one thing that you can do to make the interview procedure more reliable and get better inferences, if you must do interviews, try to structure them a little bit and make sure that everyone is getting the same questions.

David: I think that is exactly what Laszlo Bock described in Work Rules and that is what Kate book is trying to operationalize a little bit with her software called Applied. You know, the idea of let's not just have these free-for-alls, let's actually treat the interviews, as you said, give them some structure. So the interviews are in a sense generating data that you can actually then start to weigh. Is that a reasonable way to think about this?

Jason: Absolutely. There is another sense in which an unstructured interview can sometimes become a free-for-all. I was very struck by your language while you were discussing what Kahneman did with the Israeli military and how you talked about, you know, they have some idea about what competencies or skills or what have you that a person needs to do the job. Anytime you are doing any kind of screening assessing, you are trying to hire talent or even promote, you should always be going into it with those things in mind, and this is again getting back to the problem of having a clear objective. You want to keep in mind what skills or capacities or traits, what have you, a person will need to have to succeed in the role and then you should be trying

to assess those things, even if it's informal by way of an interview. The purpose of the questions you are asking in the interview are to see, are to try to see to what degree the person has the right skills, traits, capacities, whatever, to do the job. It sounds so simple, but I think during the course of an interview, it can become a free-for-all where questions sort of drift, and may be one useful perspective to keep in mind when you are thinking about what belongs in an interview and what doesn't is to imagine that you had to post all these questions, maybe they were posted online. How would you defend each and every question asked, how would you say that it relates to assessing some characteristic that they need to do the job, and if it doesn't, maybe it is not a question that belongs in there because that opens up potential to hire someone or not hire someone based on something that's not related to the skills they need for the job. So really thinking about why is this question in there, does it need to be.

David: So I suppose in some cases, like maybe if you are hiring a programmer, you can give him or her a programming task, can you code this up in Python. But for other types of employees, say, a project manager, perhaps, it may be a bit more qualitative, but that doesn't mean you can't still try to at least quantify how predictive are these questions you can put into an unstructured interview.

Jason: Structured
.
David: Structured interview, I should say.

Jason: That's perfect, exactly. When I talk about, I mean, I talk about this even in like business ethics class, like what's the ethical considerations when you are hiring and promoting and such, but I think that your screening methods should match as closely as possible what it takes to do the job. In the best case, right, you are able to onboard people and see how they do in the job and use that 90-day period for real, is this working out for both sides. Depending on how fast you are growing, how many people you are hiring, that may not be

realistic, but then stepping down from that sometimes the screening procedure is like an audition. My brother's favorite, he once worked in claims in insurance, and his favorite interview he was given a sanitized claim form and they said, here, process this and see how you do, and he loved that. As you say, some things look like this. In consulting, they do like a case interview. Sometimes when you are hiring a coder, you do whiteboard coding, although I think it is funny that they often astray from, they are often unfaithful to the way that you code or the way that you consult on the job, but the idea is to see someone do what they are going to have to do in the job. Of course, like you say, for some positions and sometimes they are higher up positions, it has to be more qualitative than that, but like you said, I don't think that the principle changes, the principle still is you have some idea in your head of what this person needs to be able to do and the reason you are talking to them and the reason you are screening them is to try to measure whether they have those capacities. I read a pretty provocative article last year about Vista Equity Partners, they are pretty successful private equity firm, and I guess they have been in the news lately because their founder just spoke at Morehouse and announced that he would pay off all the graduating seniors' college debt. But one interesting thing about Vista is that they use cognitive and personality tests to decide hiring and promotion decisions. That is not unheard of in the field but usually you think of that as happening at the lower levels. They do that right on up to the executive level and they say what they have been able to do there is to create a meritocracy and to basically the same thing, right, to arbitrage. They think they have found some diamonds in the rough that other people wouldn't have found for various reasons, right, like people had some reads on them or prejudicial judgments about them and the test showed that they had a lot to talent or that they were the right kind of person. They have been a very successful firm, and they have got it down to right. This is the ultimate in a structured assessment. It's an examination.

David: That is a really nice story to end on I think, Jason. In fact, that was just in the news yesterday, I think I heard that that he tried to recruit a more diverse workforce by using these types of analytics, but the imperative to create a more diverse workforce, that alone is a pretty huge motivation for using a more data-driven approach. There is a famous example from consulting and law firms and places like that, if I am hiring somebody, does that person pass the airport test, meaning if I am stranded in the airport for a night, would I enjoy spending my time with this person. That's always struck me as a pretty dangerous way to make hiring decisions if you actually value diversity. It just seems like a formula for making nondiverse hires. Hiring people who are like you, similar background and you'd enjoy spending time together.

Jason: Poker buddy
.
David: Exactly. So the structured approach where it's been more quantifiable,you sort of abstract away from these idiosyncrasies about like, what's the person, where did they grow up, or what are their hobbies, or what is their ethnic background, their gender, sexual orientation, those are all irrelevant for performing on the job. So if you kind of focus on the more structured questions, all that kind of noise gets filtered out, and just get reasoning from logic, you are more likely to get a diverse workforce by following a procedure like that.

Jason: Could be so. People are so confident that you can even judge the airport test from an interview, right, and the interview is such an unusual interaction where someone is on their best behavior and I just don't know how you can extrapolate from that how you are going to like working with them long term. I think there are people whose pathologies are largely turned inward, they are anxious or nervous or something like this, or there are people whose pathologies are turned outwards, and I think those people can be super confident and come off great in a short runover interviews, but they can be absolutely toxic or poisonous to

work with long run. It's very hard I think to even accurately assess the airport test from a small slice of behavior, but this is exactly why people think, one of the main reasons people think they need the interview is because they don't think that the model can assess whether you are going to like working with someone, whether they are going to fit in with the team, and perhaps it can't but I think people are highly overconfident that they can.

David: Yeah, that's an excellent point. All right, thank you very much, Jason.

Jason: We touched on some really good stuff. I had a fun conversation with you. Thanks for having me on.

David: You bet. Thanks for joining us. All right, that was Jason Dana from Yale School of Management. Thanks again, Jason, and we will see you next time in another episode of CapitalH.

The ongoing war for talent is fierce. Thanks to Jason Dana, Assistant Professor of Management and Marketing at the Yale School of Management for sharing his research insights and helping us think about what it will take to win that talent war. Join us next time as we dive into more topics and trends that focus on putting humans at the center of work.

Learn more

Visit the [Capital H podcast library](#)

Explore our [Capital H](#) blog for additional insights

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