Just Human Factors_Numbers_002

Intro [00:00:00] Opinions or points of views expressed in this podcast represent a consensus of the authors and do not necessarily represent the official position or policies of its funding. Now this is recording. RTI International Center for Forensic Sciences presents Just Science.

Dr. John Morgan [00:00:30] Hello and welcome. This is John Morgan with the Just Science Podcast, a production of the Forensic Technology Center of Excellence at RTI International. On today's show, we'll be discussing some of the challenges facing forensic science in human factors. These issues have become a rather hot topic on the national level with regard to whether forensic science is as objective as it might be. Now, of course, human factors is broader than just well whether a particular examiner is biased or not and how that's affecting their conclusions. It also encompasses things like organizational dynamics and how you recruit and train a forensic scientist. And those are all topics that the NIJ and the Forensic Technology Center of Excellence are looking at. But the issues that have gotten the most attention are cognitive bias and confirmation bias. Now, what is confirmation bias all illustrated by looking at the election of 2016? And there were a lot of things that happened during the election that were viewed very differently by the different sides. If you're a Democrat or a Republican, you might look on a particular incident from that election very differently. You'll want to see things that confirm your views and a particular event or debate, and you'll tend to ignore the things that don't confirm your biases. And things led to enormous divisions in our country. But they're part of who we are, part of who we are as human beings, and they're part of who we are as professionals. So we need to understand these issues. And there are people who make it their life as cognitive psychologists to look at these issues. Tom Busey and Heidi Eldridge will be joining us today to dive deeper into some of these human factors issues. And we're going to be looking at specifically things revolving around latent print identification. Tom Busey is a research psychologist from Indiana University, Bloomington and the associate chair of the Department of Psychological and Brain Sciences there. Welcome, Tom.

Dr. Tom Busey [00:02:22] Thank you.

Dr. John Morgan [00:02:23] Heidi Eldridge is a forensic scientist here at RTI International. She's been with us a little bit more than a year. She came to us from being a fingerprint examiner with the Las Vegas Police Department, and she's conducting a research program related to the sufficiency of latent prints here at RTI and is also a graduate student at the University of Lausanne which is one of the world leaders in forensic statistics. She is getting her PhD from the University of Lausanne. Heidi, thank you for helping out today.

Heidi Eldridge [00:02:50] Glad to be here. Thanks for having me.

Dr. John Morgan [00:02:51] Here's more from my conversation with Tom and Heidi. I was actually looking at Tom's resume earlier today, and I was very pleased because his first NIJ grant actually was approved by me when I was at NIJ and I remember it quite well, actually, because it was the first solicitation that NIJ did to even come close to looking at issues with respect to human factors or the validity of latent print examination more broadly. I see that actually in 2005. So there's over a decade worth of NIJ investment in the area represented by Tom and the work of also many, many other groups.

Dr. Tom Busey [00:03:27] You know, I was struck very early on in this process. My colleague John Vanderkolk got me involved in this research. And initially I was thinking, well, what is psychology you have to do with forensic science? We have criminal justice departments. Don't they do that stuff? What contributions could they make? And I started hanging out with the print examiners at conferences, and I would find it not during the day, but in the evening when you go out to drinks with people and they're telling you about what they do and you're talking about what you do, and you get into this conversation where people would say, How do I know if I have enough? Am I really making the right decision here? At the time I was taken back, I'm like, wait, you're the expert. How come you're asking me this question? I don't know what's enough? And it really made me realize that there are a lot of issues with respect to sufficiency, some of which can be tied to the physical print. Some of them can be tied to the experience the examiner has, the training the examiners get. And it turns out that a lot of the decisions that we end up making depend on things that I think people may not think about necessarily. Things like what's the likelihood that the detective has brought me the correct suspect? Or what is the size of the database that I use when I search for this print? Do my values that I'm using to assign my threshold, do they correspond to the values of society? So I think it has opened up a huge number of research questions that would keep many, many Ph.D. students dissertating for quite some time. Now I see. It's a huge field and some of my own colleagues in psychology are starting to get into it as well. And I'm delighted to see that because I think this is a fertile ground for research.

Dr. John Morgan [00:05:07] Yeah, it's interesting that there are different ways to answer the question of what is enough. Right. There are some folks who would argue that what you should be doing is completely removing the human from the equation. And we're going to have a completely quantitative approach to doing latent prints or other kinds of what are now qualitative disciplines. Heidi, I'd love to hear from you because a lot of your research tries to reconcile those two views. Isn't it the case?

Heidi Eldridge [00:05:33] You know, it's interesting, this question of suitability and sufficiency, you know, how much is enough to proceed with the comparison and how much is enough to render a conclusion. And, you know, there's kind of two directions you can approach that from, because on the one hand, you have the, as you say, the mechanics of the mark or the print where you're looking at how many features do I need? And we have to remember, as of latent print examiners will always say that each feature is not worth the same amount. So it's very difficult to put a number on it and say, well, you need, you know, X number of features because some helps some more than others. There's this nebulous number in the sky of if I hit this threshold, I have enough. But you're not sure how you're going to reach that threshold because you could be building up to it with pebbles or you could be building up to it with boulders, you know. And so you have this set threshold, but you're not sure how you're approaching it. But coming from the other side, there's the question of where does that threshold lie in the first place? And that's going to be affected by so many factors, some of which Tom already mentioned. You know, things like your personal values, society's values, your agency's values, even things like, you know, crime types can factor into it. You know, people don't like to talk about that because it's uncomfortable. But agencies have, you know, they have budgets, they have workloads, they have operational pressures that they have to deal with. And that comes down to making choices. And each agency has to decide how they want to address those choices. You know, where are we going to put our resources, Where do we want to stick our necks out, as it were? Where do we want to spend our time and our money? And so comes this sort of matrix that the examiner is operating within. Even on a simple question like value. You know what? What is this mark good for? Well, how does it fit into my agency

structure? How does it fit into my own risk tolerances? What I'm willing to risk if I'm wrong or what benefits we'll have if I'm right. And so a little minutia that you're looking at all of a sudden takes on a whole new perspective of this context that it's living in.

Dr. John Morgan [00:07:29] The Department of Justice describes human factors as a, quote, multidisciplinary field that examines ways in which human performance, for example, the judgments of experts can be influenced by cognitive, perceptual, organizational, social and cultural factors and other human tendencies. These factors can include examiners working double shifts so that when you're tired, you may not be as accurate or even in cluttered or disorganized work environments. Other social factors can also come into play. For example, if examiners are working on a high profile case, they may have access to case notes or information or media that imply guilt and may bias their examination. So the prototype case here that everyone points to is Brandon Mayfield. So Mayfield was a Muslim lawyer who was implicated in the Madrid bombings, I think of 2005. The Spanish police pulled a fingerprint and sent it to the FBI. The FBI ran it through AFIS. One of the top candidates was Mr. Mayfield. His print was close to the print that came over from Madrid. There's some belief and there was a major Department of Justice audit report about this that I recommend to you. It's an excellent report that talks to the entirety of the Mayfield case that maybe the examiners were biased because they knew that Mayfield was an outspoken advocate for certain causes of his Muslim clients. Therefore, they implicated him improperly. As it turned out, he was exonerated when Steve Meager actually was the FBI examiner who reexamined the prints and said quite specifically and clearly that he should have been excluded and therefore there was at least an error made. Whether the error was a result of bias or not is something that's subject to interpretation as well. But the Mayfield case certainly is the one example that people look at. Tom, your work actually even went a step back from there, didn't it focus mostly on the feature extraction process itself, how the examiner even started to look at the print.

Dr. Tom Busey [00:09:28] Yes. When we began this, we were struck by a lot of the models, a lot of the techniques that computers had used to do database searches or to do comparisons are likelihood ratios relied on features that generally were thought to be diagnostic or useful for purposes of comparison. But the models were engineered. We find these minutia, these ridge endings, the bifurcations. We have algorithms that skeletonized the print to identify those and therefore throw out a bunch of information. We thought, well, humans are outperforming computers. And that probably is still true today. But if humans are outperforming the model, then humans clearly must have access to information that the computer models don't have. So what is that information that the computers can't figure out? So rather than thinking about as an engineering solution, a sort of let's design a system and decide ahead of time what the features were, the grants really focused on first figuring out a very systematic way of extracting a large amount of eye tracking data from latent print examiners. And we have them do comparisons. Sometimes we have them do an analysis first and then a comparison. They were pretty close to a real latent print experiment. And then we took the data back to our lab and we wrote some custom software to analyze that data. And in the end we discovered that the data was incredibly rich. Now, eye tracking by itself is a proxy for what information people are using. The human visual system has a region called the fovea, which is a fairly narrow region that's only about the width of two thumbnails if you held your arms out at full length. That width is about the region where you have the highest fidelity. You can still see things in the periphery, but it's represented at a coarser resolution and so it's mainly useful for deciding where to move your eyes next and maybe getting some kind of global or holistic shape information. Once we have that information, we could then start asking questions about how can we represent that in a way that would help us do two things. One would be to try

to improve existing machine approaches, computer based approaches to matching and comparison. And the other was to try to figure out what is happening in the human visual system and the human decision process, what draws their gaze to those particular locations. We could then compare that to an analysis of the entire database to ask are those the correct locations to place their eyes? Maybe there are better regions that we don't yet know about that if we have a computer approach that could identify better regions, maybe we could suggest those to examiners and say, Hey, look over here, you aren't looking here. We know that from the eye gaze. But the computer analysis suggests that this might be a good place to look. Once you have a system that understands both the nature of the task and as well as the information that's available. You could play the two off against each other. Now, you're not going to have a perfect computer system. You're not going to have perfect examiners. Neither one is going to be sort of your ground truth. So you're going to instead try to figure out how you can improve both in kind of a bootstrapping way. So to build this system. We actually turn to information theory. And in information theory, the idea is that the features that are the rarest are the most diagnostic. So let me give you an example from face recognition. Suppose that I tell you that suspect had two eyes. Okay, that's not helpful at all because those are very common features. But if I tell you the suspect had a heart shaped mole on their left cheek, that's going to be incredibly diagnostic because that's a very rare feature. So the rare information connotes the most diagnosticity in terms of imparting information. Now, the tradeoff here is that the rarest features are also the hardest to find. So there aren't that many people that have a heart shaped mole on the left cheek.

Dr. John Morgan [00:13:24] Heidi, I'd love to know from your perspective the way in which he describes perception. Do you recognize that and how you would examine a latent print? I mean, is it possible, just from the perspective of somebody who's done it, to begin to recognize the cognitive process in that subconscious way?

Heidi Eldridge [00:13:43] Yeah, I think so. I mean, with the stuff she was describing early on about looking for the diagnostic areas and the eye tracking, and I'm just thinking about the way that I approach a new mark that I've never seen before. You know, you start out sort of globally looking at it and seeing the overall shape, what areas are clear, what areas are not clear, whether you have any anchors, which would be, you know, cores and deltas and that sort of thing visible. And then you hone in on the areas that look diagnostic. But the way that I find myself moving about the mark is I typically am going to start in an area that's relatively clear and relatively near an anchor so that I can find where I'm living. And I'm going to zoom in visually sort of focus in I'm going to be looking for features that appear to be the diagnostic and in my brain when I'm working through it. Diagnostic means something that grabs, you know, something I can hang my hat on that I see that I recognize that I would know if I saw it again. So I could have great clarity and great location that everything that I'm seeing is generic and vanilla. I'm going to move on because I don't feel like that's to help me to differentiate this mark from any other mark when I'm searching for that image. And if I find an area that's sort of rich, you know, where I'm finding lots of little features, I like that I'm going to live in that area for a while until I feel I've exhausted it. And then I'm going to go back out to that global view and try to find another area that I can zoom in on. So I could be hopping around the mark to sort of disassociated areas. But I'm trying to find tools of diagnosticity. And, you know, when we get down to defining a feature that may or may not translate to bifurcations and ridge endings, you know, that's what I'm going to be looking for first because those are sort of macro features that are easy to spot. But depending on the clarity of the image, I have a very clear image I could be wallowing around in things like edge shape that are not necessarily as reproducible, not necessarily as robust, but if you have a nice clean image,

it's information that you're taking into account and it's helping with your diagnostic facility of that area because you're saying, Wow, look at that curve there. That's really cool. I would know that if I saw that again. And so that's upping your confidence in the future.

Dr. John Morgan [00:15:47] Yes. My guess is that the fingerprint examiners who are listening would say that's where my experience comes in, because I'm able to find those rare features much more quickly and much more efficiently than somebody who's not experienced. And I don't I don't know whether that's true or not, but it certainly has some common sense value to it.

Dr. Tom Busey [00:16:03] Yeah. So that's actually, I think one of the interesting elements about comparing humans to computers. If you think about the latent print comparison process or any comparative process, the human and the computer pretty much have the same information. We're just giving them an image, a scan of a fingerprint. It's not like the computers have access to multispectral imaging or infrared or heat or whatever. The humans and the computers are essentially relying on the same information. So the only advantage that the computer is really going to have is if it has access to what you might think of as a database statistics. So how rare or common the feature is. And John, I think that's gets back to your point where that may distinguish the really expert examiners from novices, because the experts understand what features are rare and therefore diagnostic.

Dr. John Morgan [00:16:53] What do we know about how efficient examiners are based on experience or understanding the likelihood of various points of comparison?

Dr. Tom Busey [00:17:01] The work that we're actually just getting ready to resubmit for publication, we've taken a look at two computer based approaches, one based on information theory and one which is sort of formulated on a model that tries to simulate the early stages of the visual system. It tries to build models that make predictions about where experts would move their eyes to in new prints, even prints that it hasn't been trained on already. So it's not like we're trying to build a system that replaces humans. Instead, what we're trying to do is use the human data to help us understand the best way to characterize the information to the computers. What the computer approach is going to do is it's going to say, okay, this is a way of representing the visual information, and now I can use that and I can go into my database and I can say, okay, this is a feature that's common, probably unlikely to be helpful in purpose for purposes of comparison. But this is a feature that almost never shows up in the database. Therefore you might consider it to be diagnostic. Now the challenge here in both of these approaches is that we're trying to simultaneously decide both how rare a feature is because that would be very useful for purposes of comparison. But we also have to decide what a feature is. This is seems like such a simple question, like, everyone knows what a feature is, and then you think about it and it actually turns out to be very, very difficult to conceptualize what a feature is. In traditional models of fingerprints, we think about bifurcations, ridge endings as features that we would use. But for sure, examiners are using lots of other things. They're using relative position of the core of the delta. They're using the curvature of a particular region. They're taking measurements and looking at features relative to other features. So there's really a wide variety of possible candidate features. So in our approach, we are trying to be very agnostic about what a feature could be and instead we break the fingerprint down into very small patches based on where the examiner is looking. And then we can take those patches, give them to the computer model, and we can ask what is special about these regions? And we can see that a bunch of other regions the examiners didn't look at. We can say, okay, what's different about these other features that didn't get a lot of fixations from experts? And the computer can tell us, based on my analysis of sort of the

fundamental building blocks of the visual system, these are the ways that these different patches differ from each other. The actual process of defining a feature is kind of like you might think of the analogy from spoken language where we have I don't know, the average person has 30 to 60,000 words in their vocabulary, but those words are actually formed by a relatively small number of phonemes, the sort of building blocks of speech, the fricatives and vowel sounds. And we're trying to do something similar with the visual system, trying to come up with these fundamental building blocks of your visual system, and then using those to define what a feature is and therefore how rare it is.

Dr. John Morgan [00:20:20] Actually reminds me of a Neil James work. I don't know if you're familiar with what he's doing, but it's very similar in the sense that he is extracting patches of latent print, actually a fingerprint at this point, data and matching off of patches without even necessarily assuming that any particular patch has what would be considered a point of comparison within it. Reminds of one. Reminds me of the other.

Dr. Tom Busey [00:20:44] Yeah, we're probably both stealing from the same principles of that was developed. Pretty remarkable discoveries that sort of happened simultaneously. David Field at Cornell University. And then there was a group in Finland with some names I'll never be able to pronounce, so I won't try it, did what were called independent component analysis. And I can give you a sense of what's behind this because it's really kind of interesting to think about. So if you think about when you were an infant, all your learning happens in an unsupervised way. You're basically observing you're observing your mom. You're observing your caregivers. There was no what we call supervised learning, which happens in school. When you got to elementary school, you had to give an answer and your teacher would tell you if you're correct or not. But everything when you're very young and that's when the visual system is actually developing. In fact, it's developing even in utero. All of that happens due to the visual experience of the world, and the world contains structure to it. So if you think about an old analog television that is not tuned to any channel, just had the static that would be structuralist noise. There's no inherent structure in that pattern, but everything in the world has inherent structure in it. And your visual system, using very, very simple principles of heavy learning, is essentially designed to absorb that structure and wire itself up to understand the nature of the world. Break it down into simple parts that represents objects, things like balls and dogs. And so we're essentially trying to do something similar with our models and do it in such a way that it not only represents what a feature is, but it can tell us how rare or how diagnostic that feature is in an impression relative to an entire database.

Dr. John Morgan [00:22:26] A true confession here, Tom, and that is going on your website. I saw the one test you have of a random noise, a picture, basically a random noise. And you're asked, do you see a face or do you see a figure in the static? And I saw a face, but I didn't click to see what the implication was I was afraid it might say I was a crazy person or something. What is the implication that I saw based on the static?

Dr. Tom Busey [00:22:49] So this is a very interesting experiment that we did using some brain recording experiments, and we have used brain recording to look at fingerprint examiners. In that particular study, we were interested in trying to understand how the inherent kind of reverberations or noise that are in the neural networks that process visual stimuli, how they might affect your responses. So in that particular experiment, we knew that different parts of the brain process faces and process word. And what we did was we presented faces in words embedded in noise and ask you, did you see a face or did you see a word? And we could record the brain activities were active when you were doing those. And then we played a trick on our subjects, unbeknownst to them, where a third of

the trials actually presented just noise. So, of course, there's no right answer on those trials. You're like, I don't know it is something. And but we forced you to decide whether you thought you saw face or you thought you saw a word. And what we found is that on trials where you thought you saw face, the brain areas that correspond to face recognition were more active. And in the trials where what you thought you saw word the areas of the brain that were responsible for word processing were more active. And so essentially what you are seeing is that the brain, by being sometimes more active in the face area, sometimes more active in the word area, that would bias your response to toward a face or toward a word and maybe bias is wrong given all the negative connotations of bias. If you can say it gave you a preference for saying faith or a preference for saying word.

Heidi Eldridge [00:24:28] As if you were seeing both responses in the same subject. In other words, that at some points in time they would have the word preference. At some point in time they would have to face preference. And if so, they are driving the brain to be preferring one or another at a given time.

Dr. Tom Busey [00:24:42] Yeah, that's an excellent point. In fact, the reviewers ask the very same question, so you could have been a reviewer on our paper. But basically the issue is that some people that just always see face and other people that always see word. And so we did actually go into the individual subject level. And in fact, you do find that the same pattern holds up when you look at the individual subject.

Heidi Eldridge [00:25:00] Interesting.

Dr. John Morgan [00:25:01] Tom, tell me how practitioners, as you've presented this, because I know you've done that a lot and talked to a lot of practitioners and worked with a lot obviously on your research. What has been the reaction to your work?

Dr. Tom Busey [00:25:12] So I've been very impressed throughout the years at how hungry and interested examiners have been for data. I could have come into this field and discovered a group that was uninterested in change and had their fixed ways and had the sort of the handed down wisdom from their elders and would have been resistant to change. And instead, for the most part, I found exactly the opposite. People are really interested in thinking about what are the latest research on a topic. And that's been very heartening for me to come into the field and find people so open. We're always doing more experiments. I'm always looking for more subjects. So if any of the listeners would like to participate in studies, just send me an email. It's all anonymous. Most of the experiments are online and can be done in a web browser. Feel free to contact me. I think that really the next level of research for me is, I think, focusing on the dangers and the interplay of technology. The more I got into looking at the statistics of database searches. Now, this is not something that we had access to, like an AFIS terminal. So we had to sort of simulate these results. And I'm not the only one that has looked at these database problems. In fact, a lot of this work was inspired by work of ETL drawer and Jennifer Nucan. But the more that you look at the databases, the more you realize that doing large database searches is very, very dangerous. And it seems very counterintuitive because you think, well, bigger is always better. FBI has got close to a billion fingerprints. It's just going to be more likely you're going to find your suspect. And when you actually simulate that, you actually find that there is a what you might think is an optimal database size. And if you go above that database size, you run into not only diminishing returns, but you run into a situation where you're far more likely to get tripped up by a close non match or something that has sort of incidental similarity. At best, it might cause you an hour of grief trying to exclude it. At worst, it might lead to an erroneous identification. The larger the database,

the more chance of having these close non matches overwhelm the likelihood that you'll find your suspect. And so what I would recommend in these situations is when you're doing a database search, think about the size of the database for which your suspect is reasonably likely to be present. I would not run every burglary through a statewide or nationwide search. I think probably most crime is geographic and local. I used to think, why would the individual cities run their own AFIS's? Why not just have everything through the FBI? You could run this infrastructure, but it makes sense to me now that there are these local databases. But I think that the danger with large database searches is underappreciated by examiners and something to really worry about.

Heidi Eldridge [00:28:00] Yeah, that's a great point.

Dr. John Morgan [00:28:01] Very real concerns with cognitive bias. But frankly, I think a lot of the things that it examiners does to look at aspects of the case is designed to address the issue that you're talking about there. Intuitively understand the fact that if they have to look at the entire universe out there, that their accuracy is not going to be as good as if they can limit to, you know, a smaller in the possibilities of who could be matching a latent in front of them. So you actually were brought into this by a practitioner, John Vanderkolk. Tell me how you've been able to develop practitioner relationships and how that's played a role in the success of your research and any advice you can give with respect to how to improve the ability of researchers and practitioners to improve forensic science?

Dr. Tom Busey [00:28:50] It was interesting getting started with John because he wrote a letter to the chair of my department at the time, and I was a young assistant professor doing face recognition work and eyewitness testimony work. And the chair looked at this letter not knowing what to do with it. So he pawned it off on me. And I think John wrote it in response to the cases that were challenging fingerprints. I guess this maybe Judge Pollak had initially and then reversed himself, but that was the impetus in the Daubert hearings. Getting started with John was interesting because I realized that that was the first of many conversations I would have with examiners, where I would have to say similar things. And one of them was and this is very counterintuitive for an examiner, but when you're doing my experiments, I'm going to make you make errors. I'm going to put you in situations where you're not going to get the right answer. And I'm going to give you feedback, telling you didn't get the right answer and you're going to be very frustrated by that. And the reason that I'm doing that is that we don't learn anything if you get everything right. If everyone got A+ in all their classes, we wouldn't be able to figure out who needed additional math help. You're going to make errors on these tests. In fact, some of the tests are designed to be tests that really find those regions where you need additional work or where examiners as a field might not be performing as well as they could be. And so that was, I think, an initial stumbling block in ensuring anonymity in all our studies, but also giving people confidential feedback on their performance, I think helps examiners to understand why and putting them in uncomfortable situations. The second was that I think it was it was pretty clear even early on that we were going to demonstrate very dramatic differences between experts and novices. So I went into this explain it to John that we're going to run these tests and whatever we find we're going to have to publish. And if you don't like the answer, then you have to figure out how to change the world, because I'm not going to change my report. And he was fine with that. And fortunately, we never had to have any really difficult conversations because we found that, for example, with the eye tracking data, experts are much more consistent than novices are when you fix the amount of time that they have to look at impressions. And when you do brain recordings, the examiners demonstrate what appear to be qualitative differences in terms of how they're

processing the images. They appear to be processing images, fingerprints in a configurable manner, much like all of us process faces. From that perspective, it's been very rewarding to be able to document some of the strengths of examiners. The danger has always been that I don't want examiners to feel like they're super people, that these are definite differences from novices, but it doesn't mean that you're error free. It just means that you have training experience that goes beyond the average person. And it was at very least justifies putting examiners on the stand and allowing them to testify that they have real experience and expertise to bring to that testimony. But it doesn't mean that they're error free. And so that's been an interesting conversation to have with people throughout the years that really these special abilities do seem to exist, but don't let it go to your head.

Dr. John Morgan [00:31:57] So one of the things that is necessary here, I think, is to try to bridge that gap more broadly. So NIJ has directed the Forensic Technology Center to do a project to address the issue where the psychology research community has got an awful lot of knowledge and certainly has a lot of really interesting problems that are raised by forensic science. A lot of those thus far have been focused in on one very specific topic that's cognitive bias, but it's actually much, much broader than that. Meanwhile, the practitioner community is wondering what it is in many cases that psychologists are doing mucking about and forensic science and what the implications are going to be for practice. You know, we feel like this is an opportunity to have some more direct conversations there and shed some light, because you're certainly a great example of somebody from the research community who's worked very closely with practitioners. And as a result, your research has had, I think, a great deal of impact and will continue to do so. The concept is that we're going to be developing the human factors sourcebook. For those of you in the latent print community, you're familiar with the Fingerprint sourcebook, which is now a major resource for all latent print examiners, the Human Factor Sourcebook, we envision being a slightly different product than the fingerprint sourcebook in the sense that we're still trying to figure out what is important and what can be said about human factors in forensic practice. And so what we're going to be doing is getting roughly a half a dozen researchers, together with a similar number of practitioners, to look at very specific kinds of topics in psychology and their intersection with forensic practice. We're going to be working on that over the course of the next months to put together that group of folks. And Tom is very graciously agreed to lead the effort. On behalf of the other psychologists, thank you very much for working with us on that, Tom. Sure. I'd love it if you could just give me your views with respect to kind of the scope of psychology and the implications of research within the forensic science community. What are the elements of psychology that you think might be useful in the long term to improvements in forensic practice?

Dr. Tom Busey [00:34:15] I'm thinking one of the things I'm hoping that will come out for me personally of this project, I've really started to think about getting into the heads of the day to day activities of a latent print examiner. What are the stresses? Is it the backlog? Is this the repetitive nature of the work? Is it eyestrain? Is it fatigue? Is it about making a mistake or worry that you're not performing as well as other people? I really feel like the solutions that we come up with to address those issues are very much a kind of a management solution. Not in the sense of we need to tell managers how to better manage their employees. But I think what an ideal lab environment would be like that. I would love to see, for example, a ongoing seminar series, individuals present manuscripts from the literature to the group and lead a discussion about the strengths and weaknesses of that paper. And having that be a weekly or monthly activity to get people in this sort of in-service mind thinking about how do I continually improve my skills and avoid being sort of stuck in a rut. So I think that latent print examiner centric view is one that I think would be

very fruitful when we think about, the danger is that psychologist will say, well, I did this in the lab and I think it's going to work. And so now everyone needs to do this, and that doesn't seem to be that top down approach, doesn't seem to be very successful. So I would love to hear from examiners about what it is that stresses them out at work and what can psychologists do? You know, I try to I tell people when they hear I'm a psychologist, I will pretend to listen. But I think in this case, even though I'm not a clinical psychologist, I think that I do hang out with a lot of clinical psychologists in my department. And I perfectly give me the skills, I think, to help me listen to latent print examiners when they talk about their jobs. I would hope that the first step for the broader psychology community would be to understand the nature of the task, not even just prints. What are you going to decide? But what are you hearing from management about the number of erroneous IDs or erroneous exclusions, for example, that you've been making? Or how do you tell someone that they're making too many inconclusives and they're not really doing their job? Or maybe the amount of inconclusives they are doing is reflective of their current capacities. And are there ways that we can do some training to improve the rate that they can actually make a decision? So my hope is that this is not a scary process for examiners that that we can go in, make some recommendations or suggestions that would improve the quality of life for them in their jobs, which I think ultimately will then percolate down to better forensics overall, that better decisions get made because people are liking what they're doing and really applying themselves on a day to day basis.

Dr. John Morgan [00:37:15] Tell me, Heidi, what do you think are the top 2 or 3 issues that implicate human factors are in the forensic laboratory yourself? What's your view of it?

Heidi Eldridge [00:37:25] I think that a couple of the big ones really have to do with the shifting culture that Tom was just describing a little bit. I was just thinking as he was talking about what he would like to see happen in the lab, that the time is really ripe for that right now. And I think we're really becoming a more open place, vocal practitioners and managers where we can have those sorts of discussions about what what's a reasonable baseline of those exclusions, where could there be training done that would improve people? And it's being done in a more constructive way now because even as few as 8 or 10 years ago, I think the laboratory culture generally was a much more punitive place. And so there was this very much a fear culture that if somebody finds out I made a mistake, I'm going to lose my job. That really colors the way you do your job when you're always in fear of losing. And I think now that there have been enough changes in the way discussions are happening within the discipline, that people are more open to discussing these things in a non-threatening, a non-threatening way. How can we improve our baseline? How can we identify people who need additional help? How can we get them that help? So I think that that is one area that's a huge even the fact that we're willing to have these discussions and people aren't so afraid anymore. Another one I think has to do with again, I touched on this earlier, but just identifying agency cultures, not that there are good or bad thing inherently, but I think that when people are trying to make decisions, it's helpful for them to be cognizant of the culture that they're operating within and how that culture may be influencing the decisions. For better or worse, they're still going to make the decisions, but it's helpful to have a clear vision of why they're making those decisions in the way they are. And then the other two areas that I think are really open for improvement right now are both fitness for duty and selection training of new examiners. So in other words, first of all, identifying when people aren't at their top and should maybe not be making critical decisions, whether that's because of fatigue or sickness or, you know, just having a personally bad day, but being able to identify that, be able to self identify that and feel like you can go to your manager and say, you know what, I shouldn't be concerned today, let me work on an administrative task or something like that. And then also just identifying

what skill sets make other examiners successful and try to identify those at the recruitment stage so that we can get the right people and get them the right training to help them succeed.

Dr. Tom Busey [00:39:56] So one of the issues that that has come across in terms of the things that kind of bugs me, I guess, about the testimony that latent print examiners give and I'm could be totally off base here, but I feel like if we're going to be open and honest, examiners often rely on when they make a conclusion, they'll say, well, based on my training experience, I've never seen to impressions that came from different sources that had as much detail in agreement or some phrase that sort of similar to that. And the language may be changing. I don't know how typical that is anymore, but one of the things that I would really like to emphasize is the role of feedback in training, the feedback that you get when you say, I've looked at X many thousands of prints in my career and I've never seen two impressions that are from different sources that have this much detail and agreement. The question is, if that occurred, would you know it? Because it's my impression that most of the comparisons that the examiner is doing, most of the images that they're looking at, are ones where you may not know the ground truth. They're part of casework. And the whole job of casework is to determine ground truth or determine if these may have come from the same person. It makes me a little nervous that a lot of the experience that the examiners are relying on are not sort of ground truth experience. And so I would recommend that and I know some of these resources exist, but I think it's really important that on a fairly regular basis, the examiners have access to impressions that really are these close calls. So I know there exist close call databases. We've developed a tool that not only gives you feedback about whether two impressions came from the same source, but they also give you feedback right down to the individual ridge that you're trying to trace and if people would like access to that data testing, I'd be happy to share that. I really feel like feedback is something that will help us establish thresholds would help individual examiners decide if their current level of sufficiency is consistent with the values of society, really would help on an ongoing way to address the sufficiency issue, whether your skills need to be tuned or are there particular kinds of prints that you'd come to the wrong conclusion on. So I want to stress the role of ground truth feedback as the best way to improve your skills. Whether you find out from situations where members of your lab create simulated latent prints, where they know the ground truth and then have exercises that they develop or go to online resources, but like to stress the role of feedback and really improving examiners skills.

Heidi Eldridge [00:42:42] Yeah, I think that's an excellent point. Tom, I'm really glad you brought that up, that more emphasis needs to be given on sort of continued training, keeping up with the skills and a lot of laboratories send people to additional courses, learn about statistics and learn about mythology and learning how to search palms that don't see much ongoing training with the sort of feedback reinforcement you're talking about, it's incredibly valuable. The other thing that I have seen that some laboratories do is they'll have occasionally what they call recalibration exercises, where they'll get the entire unit together and they'll do some of these sorts of things together. You know, everybody compare the same marks or everybody take the same marks and analyze them let's see what really should they mark, or just some sort of basic skills. But by doing them all as a unit, what I mean, they do them individually but then meet as a unit to discuss them. Then it really helps to reorient everybody in that unit to being kind of on the same page about what they do consider suitable for what they consider to be too degraded to count as a reliable feature and that sort of thing. And I find those kinds of things are useful both for keeping the skills sharp and for getting that feedback from your peers and sort of keeping cohesive in your agency culture of what you think is appropriate for this occasion.

Dr. Tom Busey [00:43:56] Yeah, I like it. And I think that a fear in that I think is that if we do everything together, somebody is going to discover I'm not as good as somebody else. And if I can use a basketball metaphor, you're always going to have five guys on the court and they're going to be differences among the players. Players bring different strengths. Some people are better on defense, some people are better on offense. Everybody has a different skill set that they bring. But you can still play five players. You're still going to sub in people at different times and there always is going me an opportunity to improve. It's okay if you're not the best identifier. Maybe you're really good at excluding prints very quickly, so that might be the opportunity to discover where people excel. Maybe some people are really good at in Hydro prints, but other people are good at Black Powder prints. Those exercises should not be viewed as threatening and I really stress that management should not use them to penalize or give different raises or whatever to examiners, it really should be an opportunity to share everyone's strengths. And if that sounds like a Kumbaya statement, so be it.

Heidi Eldridge [00:44:58] No, I agree with you. And I would even go a step further and say that if it's done correctly, it's an incredibly valuable tool for the manager because now that you know who you're good hydro person is and who your good excluder is, you know, not only can I help you make decisions on who should be doing those sorts of cases, but it also gives you a built in trainer. You know you can save the rest of your unit. Hey, everybody, go see Joe. He's great. in hydro, he can help you get better in hydro, you know. And that way you're utilizing your strengths for everybody.

Dr. John Morgan [00:45:28] Well, I appreciate the idea of a kumbaya moment, and I think that makes a good wrap up for our podcast today. A lot of really great substance and a lot of really great discussion. I know everyone very much enjoyed hearing from you, Heidi, and from you, Tom. And excellent time just to spend on this snowy afternoon talking about human factors and forensic science. And I'm looking forward to working with you on The Human Factor Sourcebook and seeing a really great product that I hope will be of great use to the community.

Dr. Tom Busey [00:45:59] Yeah, it's really my pleasure.

Heidi Eldridge [00:46:00] Yeah, it was great.

Dr. John Morgan [00:46:01] Thank you very much. I appreciate the idea of closing on a kumbaya moment and we certainly do need additional tools and training in this area. As mentioned earlier, the Forensic Technology Center of Excellence is working to develop a sourcebook that will cover this topic as well as much, much more going into some of these issues outside of how an examiner works and looking at more organizational and personnel issues as well to try to help us not just with standard operating procedure development, but other issues in the forensic laboratory related to human factors. We're also developing a leadership training that will serve as a way of helping us learn how to cooperate and organize among each other and demonstrate leadership principles that are relevant to forensic science practice. I want to thank Tom and Heidi for their participation in this podcast, if you'd like to learn more about their work you can visit our website at www.forensiced.org/justicesciencepodcast. And next week this is what we're going to be talking about.

Preview of Next Episode [00:47:06] Let me start by taking this one slice at a time.

Preview of Next Episode [00:47:08] Decanter and (indiscernible) check.

Preview of Next Episode [00:47:10] There's two data set. People can test their hypotheses or models they can test using that.

Preview of Next Episode [00:47:14] You've also been developing picture interpretation tools that are highly regarded, especially if you get into complex mixtures.

Preview of Next Episode [00:47:20] Desmond's group did develop two tools that we have placed online provide the probability distribution on the number of contributors to a sample.

Preview of Next Episode [00:47:29] Help with procedural development and processing of cases in the crime laboratory.

Preview of Next Episode [00:47:34] And I think that is the path forward.