Rationally Speaking #219: Jason Collins on "A skeptical take on behavioral economics"

Julia:          Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef, and my guest today is Jason Collins.

                Jason is an economist based in Sydney. He's the leader of the data science team for a financial services regulator and blogs at jasoncollins.blog, which is one of my new favorite blogs. It's got very in-depth and interesting reviews of books and all sorts of areas in economics. And commentary on different aspects of the important debates in behavior economics and evolutionary economics in related fields.

                Jason, I'm excited to have you on Rationally Speaking, finally.

Jason:          Yeah, really pleased to join you. I'm a big, big fan of the podcast.

Julia:          Oh, yay. First off, one thing I wanted to ask you is whether you consider yourself a behavioral economist, which — not that it matters that much, it's just a label. But I'm curious whether the criticisms you are likely going to make in this podcast, of behavioral economics, are coming from an insider or an outsider. Just kind of changes the tone.

Jason:          I think I'll call myself an insider, because deep down I am a big fan of behavior economics. It's just a case of, I think, that behavior economics or, as I prefer to call it a lot of the time, behavioral science, could be so much more.

Julia:          Why don't we discuss that distinction before we dive in? Why do you prefer the term "behavioral science" instead of "behavioral economics?"

Jason:          Well, I think in some cases behavioral economics is the right term, where you're bringing psychology into analysis of economic problems, economic decision-making. So that's probably the right term there.

                But the label "behavioral economics" actually gets applied to a much broader area that includes a lot of pure social or cognitive psychology, a lot of other decision-making fields. And I think to give credit to the people working in those fields — they're not economists, they're typically not working on purely economic problems, and we really should give it a label that captures the fact that they are adding interesting inputs to how we think about human decision-making.

Julia:          Is the label "behavioral economics" just a historical accident, because it started with people applying the stuff to economic models? Or does it persist because it's somehow more respectable or something like that?

Jason:          I think a bit of both there. Because at the beginning, if you go back to the work of Kahneman and Tversky, they really were setting themselves up against the work of economics, so the foundations of it really come from an economic angle.

                Now it's probably a smaller case of a marketing term. When I'm talking to people and trying to get their interest or attention, quite often I'll use the word

"behavioral economics" to get a sign of recognition straight away, instead of the blank looks when you say behavioral science.

Julia:     One of your somewhat-recent lectures was titled "Please, not another bias," and in it you referred to the Wikipedia page on cognitive biases, and you said, "There are not 165 biases." Can you explain why you say that?

Jason:     Sure, it comes back to that -in some ways- foundation of behavioral economics. In the initial work of Kahneman and Tversky, they - and in fact, some of their predecessors - looked at some of the axioms of neoclassical economics, particularly around the idea of people maximizing expected utility. So, take a lottery - we're choosing whether or not we're going to accept that lottery, and expected utility theory tells us what sort of axioms there are around the person's choice and how that'll weight the outcomes according to the utility function. And then look at the probabilities, do a bit of a sum and get the answer.

A lot of that work with Kahneman and Tversky was pulling that apart and basically showing the first holes in that model, saying, "Look, people don't actually decide like this. When we give them a choice under uncertainty, there's systematic deviations from this expected utility model."

That's a great start, and what's really happened over the last forty-odd years now, is we've seen a growing number of deviations from that model - an endless array of points where people simply don't confirm to the way this economic straw man started.

So now we have this kind of a mess, and you look at that page... The last time I counted, there was about 160-odd [biases] on it. I think it's even grown since then as people just keep simply going, "Here, here's another way in which people don't conform." But the problem with it is that we're not really in a world where that economic picture is actually the accurate one, so it is simply creating more and more deviations from the wrong model.

In that talk I use the example from astronomy, so — you go back to 1500, I think I've got the right year there, and we had this model of the earth at the center of the solar system, and the sun and the planets moving around the earth. That's a decent model, but then of course you start to observe all these deviations from that model. So we see that Venus only appears in the morning or the evening sky. A lot of the other planets such as Jupiter will track across the sky, but it won't just keep going as you'd expect; it was in reverse direction. So people started creating what effectively were called epicycles, different patterns of movement of these planets, and creating an incredibly complex model of the solar system.

Then of course, Copernicus comes along, points out that the sun is actually at the center of the solar system, the stars and the planets... Sorry, the earth and the other planets all went around the sun. Suddenly you've got a much simpler model.

And I think that in some ways, the world of behavioral economics is a little bit like that pre-Copernican world. We're spending all that time talking about deviations in what turns out to be the wrong model.

Julia:      Haven't there been attempts to propose alternate models? Like would Kahneman's prospect theory, for which he won the Nobel Prize, would that count as an alternate model?

Jason:      It's starting to move there, but Prospect Theory is probably a case in point where effectively their starting point is the expected utility, and then they add in some different elements of it. So the utility function, instead of just being... diminishing to a great extent as it gets larger, there's reference-dependence and loss-aversion in it. Instead of just weighting directly from the probabilities, there's first this function of transforming the probabilities.

            But in the end it's built on that neoclassical model, and you end up with a model that is no more realistic a statement of how human decision-making works... It's a descriptive model, rather than a model which actually says, "This is the way that humans decide." Which leaves you in what [Berg and Gigerenzer] called "as if" behavioral economics.  So behavioral economics, in the same way as neoclassical economics, just ends up talking about humans behaving "as if," rather than going "What's the real decision-making function here?"

            So yeah, prospect theory, you're right. It's kind of starting to try to pull together a few things. But you've still got a lot of other cognitive biases that are outside of prospect theory, and even prospect theory itself is this cobbled-together thing which has a lot of the same flaws, in fact, that neoclassical economic approach does.

Julia:      All right. So the unifying theory in the case of the epicycles was heliocentrism, and elliptical orbits instead of circular orbits. What is the unifying theory in the case of cognitive biases?

Jason:      Well, if you had asked me this question several years ago, I would have jumped straight into an argument that evolutionary biology would be that unifying theory.

            I still think there's a grain of truth [there], and in fact I think any unifying theory will have to have an evolutionary component. I suppose what I'm a little bit less sure about now is just how clean that model alternately looks.

            Even look inside evolutionary psychology, and it has the modular theory of the mind, where all these different modules of the brain can be triggered in different circumstances. And it's not a particularly clean way of developing a new model. And so perhaps we may end up simply here in some ways that rational [actor] model that we have from economics might just be our benchmark for a very long time.

Julia:      Ah, I mean to be fair, human psychology and behavior and societies are much messier systems than astronomy, so...

Jason:      Yeah, exactly.

Julia:      It's a bit of a high standard to which to hold ourselves.

Jason:    Yeah, in some ways to call for a really clean, unified model, I guess it's... We love clean, mathematically beautiful models, and perhaps we're looking for something that actually isn't there.

Julia:    When I asked about the unified theory a moment ago, you said "evolutionary biology." But is there a way to state it as more of a model instead of a topic? Like... maybe that's a tall order, but something like "humans make decisions in a way that maximizes... they follow heuristics that maximize the genetic fitness of their ancestors?" Or something?

Jason:    Yeah, I think that's probably right. Basically, humans pursue a set of proximate objectives that, in the environments in which they evolve, would have met that ultimate objective of increasing their fitness, [via] survival and reproduction. So the parts which evolutionary biology would add would be around, "What are the objectives that people are pursuing?"

          ... In economics, we typically talk about maximizing a basket of consumption or something of that nature, yet evolutionary biology sort of points out, "Well what is it going to be?" Things that lead to survival and reproduction. And then of course that might lead us to go, "Okay, well there's threat avoidance, there's disease avoidance, there's desires for certain foods," and you start to build from there.

          But even then, that's only a part of it, because let's suppose that gives you the idea of our objectives, but then what's next? How do we think about the shape of... using the word utility function, for lack of a better word. How then, do people make choices over those different options?

Julia:    Yeah, so I actually came into this podcast all ready to disagree with you about how much the evolutionary lens adds, but it's possible that we won't disagree with each other as much as I expected. But let's throw a couple of examples out anyway and see what you think of them.

          One cognitive bias that didn't seem, to me, to fit the evolutionary framework is the conjunction fallacy: the fact that in some cases, we assign higher probability to A and B than we do to A alone. The classic demonstration of this being, someone says, "Imagine a person named Linda. She's left-wing, she attends a lot of protests, et cetera," and then they ask "What is the probability that Linda is a bank teller? And also, what is the probability that Linda is a bank teller and a feminist?"

          People tend to think the latter, the bank teller and feminist, is more likely than the former. Which is not possible, because being a bank teller and a feminist is a subset of being a bank teller, so it can't have higher probability.

          That kind of bias, and there are others I put in the same category, doesn't seem related to humans maximizing their genetic fitness, it just seems like we're bad at reasoning about probabilities.  I could give other examples, but how does that kind of bias fit an evolutionary model?

Jason:    I think you're right on one hand. Sometimes a conclusion may just be we're bad at some things which we've never encountered before in our evolutionary past.

But I think a different way to think about it is, instead, start, "Okay, what are the sorts of cognitive tools that evolution might have shaped? What sort of heuristics might it have given us?"

This leads us, in fact, to territory where Kahneman and Tversky have a lot of debates with Gerd Gigerenzer around the Linda problem.

And there it may be a case of... they assume that all information is relevant. Tying that back to how that might have evolved, there's a few steps that need to be filled there. Then you go, "Okay, if you have someone who's [under the assumption] that all the information you're given is relevant," then you can start to see, "Okay, well here's where people may start to make this leap and make this mistake."

I have to say, though, there's a set of papers in 1994 where Kahneman and Tversky and Gigerenzer actually debate this point back and forth. It's really quite an entertaining debate, and I actually kind of leaned a little bit with Kahneman and Tversky on the fact that there is really something here that's tough to reconcile...

Julia:    Yeah, I keep noticing this weird feature of debates over whether humans are irrational, or to what extent cognitive biases are really biases, and the debate between Kahneman and Gigerenzer is a great example of that. Where one side, behavioral economics or cognitive science will say, "Look, people's choices fail to maximize their own utility, so they are irrational."

And then the other side, the evolutionary side, will chime in and say, "Ah, but people are making choices with heuristics that maximize the genetic fitness of our ancestors." Or, "But we got the right answer in the environments we evolved in, and so they are rational."

And as far as I can tell, those two perspectives are totally consistent with each other. One side is saying, "Our choices are suboptimal in the modern world for us as individuals" and the other side is saying, "Our choices were optimal for our genes in the ancestral environment." But these two sides seem to think they disagree with each other.

Whereas this whole debate seems to me like two sides talking past each other, because they're using different definitions of rationality. Or they're arguing about "How much should we wag our finger at humans" or something, which doesn't seem like the most important question. What's your take?

Jason:    I'm with you. It does seem like one of those debates where they really do agree with each other. Looking at Gigerenzer's work, in fact, so much of it is around how people fail to understand probabilities, and then he of course proposes alternative ways of presenting those choices to people, typically using frequencies instead of numerical probabilities, and you get better decision-making. And that, if anything, is a perfect example of a nudge, which is exactly the same thing he'll often rail against.

So for me, it really is a case of an argument more around framing than real substance. Although, again, going back to that 1994 paper, there are probably some points where they state that they disagree, so Gigerenzer will probably fight

about more, "Where exactly is that line?" He'd move it much closer to the "More of these decisions are consistent, or error free," than perhaps others would say.

Julia:     Hmm

Jason:     In fact, I should say: the big bit where I have a lot of sympathy for Gigerenzer, in fact, I'm a big fan of his work, is this idea that, ultimately, all of our decisions, they don't conform to this model of rational decision-making. So they often involve us excluding information from our decisions, or deliberately making biased decisions.

His point being that they're designed to operate in the real world. They're not designed to conform with these tenets that we've decided on, on how people should decide. And so when you start going, "Okay, let's put someone in an uncertain, unstable environment, and with constrained computational power: what's a good decision-making strategy?"... suddenly some of these biased approaches have a lot of value.

And now you're seeing in work from people such as Tom Griffiths -big, big fan of all of his work- which is still saying, "Well actually, if you look at computer science and the way they solve problems, you're actually seeing the same pattern. That adopting biased solutions, things that work well in the real world, for the problems we're facing and the resources we have at hand to solve them.

Julia:     Okay. Well, one key point that we sort of skipped over was: What are some examples of cognitive biases that you think are better-explained by the evolutionary lens, by the lens of humans trying to maximize their genetic fitness, than by the standard behavioral econ approach?

Jason:     There's one, the availability heuristic - there, of course, that's the idea that people judge probability based on simply how available recent examples are in their mind. In a world where you've got media confronting ever small disaster and the like, it's a fairly flawed tool. [If there's been] a recent terrorist incident, people will judge the probability of terrorist instances to be much higher than they actually are. Or shark attacks, anything like that.

This is... I suppose this is the sort of thing where you go, "Okay, in the modern world it backfires." But then go back into an uncertain environment and you're then faced with the same problem, and you're getting a simple, small sample of these events.

Here, you end up with a scenario where actually the availability heuristic actually becomes an optimal approach. In fact, I recall Tom Griffiths, when he spoke to you in one of these previous podcasts, actually talked about a paper where they worked with... I forget if it was a co-author or the person who actually did that work, but sort of this idea that, in certain environments, that's the way to go. It's not a case of really trying to get this representative sample, it's accepting that you're only going to get a drip feed of certain examples, and you're going to make judgements based on them.

Julia:     Right, right. You know, often when I read evolutionary psychologists or evolutionary economists talk about biases secretly being rational, they're talking

about biases that serve kind of a signaling purpose, like when people are overconfident or they overestimate their... whatever good traits they have.

Jason:      Yes.

Julia:      Or the way they reach conclusions about political or economic or other ideological facts is biased by their preconceptions or their tribal affiliation or something like that.

And the evolutionary psychologists say: "Actually this is rational. Because our goal is not to reach the truth in these cases, it's to signal various things to other people about our genetic fitness, or about our loyalty to the tribe or things like that."

Would you also agree with that explanation? Would you call those behaviors rational for the individual, or just rational for maximizing our ancestors genetic fitness?

Jason:      Yeah. I think that's probably the core question for a lot of these ideas. It's one thing to say, "Okay, there's a behavior that some time in our evolutionary past maximized whatever objectives we're trying to achieve," but how does that actually line up now? The reality is is most of us don't walk around seeking to maximize the number of genetic offspring that we have. We actually have other objectives, and yet that's what a lot of these proximate objectives we pursue are about.

Julia:      Right.

Jason:      But I think you're right there. There's a whole idea of, I guess, costly signaling theory is actually one of the really promising areas — we can go, "Okay, well why are people behaving in this way?" Suddenly it pulls you into areas where you go, "Okay, here's some evolutionary insight as to what's going on."

Costly signaling is around pointing out that you're higher quality. But it's relative to other people, and the neoclassical model doesn't have that relativity, you start to see that in all the behavioral literature. So, it starts to give a foundation for, "Well, why do you want to be relatively better?" Or you're competing for, say, scarce mates, which there's a constrained supply of them, or other resources, which are similarly scarce. Nowadays it's scarce properties, scarce schools and the like. But you enter into a real relative competition, and you need to either signal the right quality to win them, or alternatively win that relative competition to get access to them.

Julia:      Is there any other kind of mismatch between the evolutionary environment and the world that we live in today that might make these evolved features of our cognition less advantageous for us now as individuals than they were for our genes?

Jason:      Yeah, well I actually really like the argument that Geoffrey Miller makes in his book 'Spent'. It's basically this idea that in a modern consumerist society, we're picking up all of these different ways of signaling our quality, which are actually pretty crappy ways of signaling our quality. You go buy an iPhone or something

like that and it's a pretty conformist step, it's not signaling a lot of wealth or status, and yet a lot of people probably grab them thinking exactly that. Instead, when the traits we're seeking to signal are things such as intelligence or creativity or how caring we are, there's probably far better ways of doing it than engaging in some consumerist spending. Instead it could be ... whether it's around storytelling or making objects or engaging person to person, there's probably better, more effective signals that we could be using, rather than looking for the latest little present that we can buy.

Julia:     Yeah.

Jason:    It's quite simply a case of marketers and the like. They're looking for every angle they can exploit to get us to buy their products.

Julia:     Right.

Jason:    We, of course, aren't particularly well adapted to that degree of onslaught. The same thing happens in food of course. Most food nowadays, it's been developed increasingly with] tweaks towards something that's hugely appetizing for us. We're not particularly ... well, a lot of people aren't particularly good at defending themselves against that temptation.

Julia:     Right. You wrote an interesting blog post just a couple months ago about cognitive biases, and you argued that naming biases like the "availability" bias or "loss aversion" or whatnot, that gives us the illusion of having explained the corresponding behaviors, but actually we haven't explained anything. Can you elaborate on that?

Jason:    Yes. Naming something does give you a little bit of knowledge. Naming, say, loss aversion does at least give you the bit of knowledge that people may weight losses more than they weight gains.

But just knowing the name of it is a pretty weak or thin form of knowledge. I mean, think about all the other questions and points you might want to know about it. Why does loss aversion exist? What context is it most powerful in? How does it operate? So, what are the heuristics that we use that ultimately lead to this loss aversion being shown?

It's only through being able to think about, and answer a lot of questions like that, that you get to a point where you can go, "Okay, well, in what contexts should we expect [this bias], and how can we design so people don't fall for it if it's a problem in their decision making?"

I'd even go as far as saying that sometimes knowing the name of something can actually cloak the reality of what's going on underneath. The one I think about a lot is overconfidence. In fact, you had a great conversation with Don Moore on this very point previously, where you can break it down into at least three different conceptions of overconfidence: overplacement, the idea that you're better than others; overestimation, so the belief that you are simply better than you actually are; and overprecision, the idea that you're, I guess, more accurate.

Julia:     Right

Jason: So, you might give a narrower 90% confidence interval than you really should.

Then you think about each of those three, and it's not a case of which we're pervasively over placing, over estimating and over precise. We tend to over estimate for hard, difficult tasks, but we tend to under estimate when it's easy. Conversely, we tend to over place on the easy tasks and under place on the harder ones.

So, you actually have this pattern where you're actually getting ... if you use this word, over confidence, you're actually over and under confident, depending on exactly what the task is. It's only by getting past the name and thinking about, "Well, what are the different ways we can describe this?" Then in turn getting to that next layer of questions. Well, in what context does it actually apply that we get to a useful understanding, as opposed to just a label that we can throw at poor decisions?

Julia: Right. Do you, in fact ... I'm curious to what extent do you think people are actually mistaken about the amount of explanatory power these cognitive bias names have? On the one hand, maybe they're just saying, "The term loss aversion is pointing at a pattern that exists in the world, so we can make predictions now that we know that this pattern is there. We're not claiming to have an explanation for why humans are set up such that they are loss averse, we're just pointing out the pattern."

That would be the non mistaken world. The more mistaken world would be one in which people are just kind of confused. It's sort of like saying, "Well, the reason that this ingredient makes you sleepy is because it possesses a soporific quality," where soporific just means a thing that makes you sleepy.

So, it's like it has the ring of an explanation, but it doesn't actually really explain anything.

I've heard some convincing arguments that aspects of neuroscience are like this. They'll point to a part of the brain that lights up when someone's doing a certain kind of cognitive behavior, and say, "Ah, the reason that we do this behavior is because this part of the brain is active." They're not really explaining anything, which is the situation that you think we're in with respect to cognitive biases.

Jason: A bit of both. On loss aversion, quite often you'll hear people say loss aversion, well, it's a function of prospect theory. But in turn, prospect theory is just a cobbling together of a set of biases such as loss aversion. So, you actually see that a bit here. People fall for it.

But on that first point around simply being able to label it and say, "Okay, it's a general pattern we can make predictions on"... I think people over estimate the extent to which you can do that. Across the whole literature on loss aversion and the endowment effect, which often gets wrapped into it, there's actually a lot of, I guess, context dependence, or at least results where you don't see that pattern.

It's at a point now where there's been recently some papers ... or I think this debates been going for a little while now, but just how pervasive is loss aversion really? You see this again and again in different pieces of literature. A

confirmation bias I was recently reading, I guess an argument there that it's not quite as pervasive and useful as people think. I think it's probably both layers of problem there.

Julia: Yeah. I mean, another reason to think that there might not be 165 cognitive biases is just that some of them might not be real. They might be the result of P-hacking or publication bias or some other methodological problem. So, I'm asking you to speculate here — If you had to guess, which cognitive biases do you think are most likely to be real, and which are most likely to not be real?

Jason: Well, in some ways, it comes back to my initial thing. In some ways, a lot of these biases don't exist because we're using the wrong reference point. It's not so much a case of saying these heuristics and biases don't exist, but it's probably going, "If you think about them in a different way, well, would you still describe them in that way?"

For instance, I think the availability heuristic representative, like, in fact, most of the early work of Kahneman and Tversky... I think it's pretty robust, most of it. It's highly replicated. But ultimately, if you adopt a different model of decision making and you go, "Okay, our benchmark is no longer this model of rationality that we've came from economics. Instead, our benchmark is real world decision making." Suddenly, you may not be thinking about availability heuristics anymore, you may think about this as some different cognitive tool, which may or may not work in other environments.

So, I think some of it won't simply survive a more theoretical reframing. On that list of the Wikipedia list, I think there's just a lot there that are going to disappear. Not such a case of P-hacking, but it's the case of having thousands of researchers around the world running endless experiments and turning up neat things, and all it takes is a single paper and you've got something you can throw onto the Wikipedia page yourself. A lot of it is going to fall away from that angle.

Julia: Got it.

Jason: The other funny thing with a lot of them is they're even in different directions to each other.

Julia: How so?

Jason: Well, recently had someone talking about status quo bias and action bias. So, I don't think action bias is actually on the Wikipedia page, but it gets talked about a lot, this urge that people feel the need to take action.

But of course then you've got a status quo bias, which is people want to stick to the status quo. Well, which is which? In what environments are they active - is one of them active in some contexts and the other in other contexts? Or in the end, have we just simply got two results and labeled each of the two results as opposite effects?

Julia: Right. It's like having two opposing idioms that together encompass the whole space, like "opposites attract" and then "birds of a feather flock together." And you can pull out whichever one, to explain whatever you see.

| | |
|---|---|
| Jason: | Indeed, indeed. You see a poor decision from a CEO, and if they chose to do something, well, they were over confident - but if they didn't make the call they should have, they were loss averse. |
| Julia: | Right. |
| Jason: | There's sort of a bias for every situation. |
| Julia: | Interesting. Shifting focus a little bit now, you had an article just earlier this month in which you express concern about whether the pendulum has swung too far in the direction of behavioral economics or behavioral science in policy making, as opposed to economics in policy making. What is the problem there that you're concerned about? |
| Jason: | It's simply a case of focus and investment of energy. Right now, behavioral economics is a pretty sexy area. It has been for a few years now. We're seeing a real proliferation of behavioral ... well, generally branded as behavioral economics or behavioral insight teams. They're sort of the main names given to them. |
| | But at the same time, we have a lot of these organizations that are really investing in building that capability, and they either don't have or have much smaller economics capabilities. I suppose my complaint was ... really around the question of the sorts of problems these organizations are facing — how much of them are actually a pure economic problem, or a behavioral problem? And are we leaving a really important part of the toolkit on the bench? |
| | I should say ... my argument is a little bit broader than that. I think there's probably just a case that for these sorts of problems we should be bringing in highly multidisciplinary toolkits to try and solve the problem. Many cases that won't be just behavioral economics, or economics that might be from other areas, it could be from varying areas of psychology we haven't exploited yet — it might be anthropology, it might be other areas. |
| | Now, to put a bit of flesh, though, on why I think it's a problem, is that many of the issues that these behavioral teams are getting thrown at are fundamentally economic problems. It's quite often the case where the problems they're trying to deal with are effectively, incentive problems. So people are, say, making poor decisions in a business — and yet the real reason they're making these poor decisions is basically because they're being paid to make those decisions. They're selling products to customers that they shouldn't be selling to, but why are they doing that? Because they're paid to sell more products. |
| | Then you've got behavioral teams that are being sent in to try and, in some ways, push against the tide. And they're sometimes not being equipped or not being given license to go, "Well, what's the real problem we have here? If we're going to go to a good solution, what should we really do?" |
| | We see a similar pattern I think in just policy making today. It may be partly a question of political feasibility, or partly a question of people having the confidence and courage to put forward options where the behavioral option is the easy one. So putting some, I guess, notes or some comparisons onto their power |

bill with their neighbors so that they might reduce power, is a lot more palatable and easier than going, "Okay, well, could we price, say, price carbon properly to get that reduction, or increase the tax?"

Julia:        Oh, interesting.

Jason:        So, quite often I think the more powerful solution isn't getting a look in.

Julia:        So, the behavioral insight style solutions are both sexier and also easier, which is a dangerous combo.

Jason:        Indeed, indeed. Maybe, yeah, it's even a case of availability. There's always endless presentations now around ... and they are successes, we shouldn't play that down, but when people tweak a letter to increase tax collections and the like — really cool ideas, but the less sexy sort of basic economics just don't get the same degree of play nowadays.

Julia:        Would all of these kind of behavioral insight style interventions fall under the heading of nudges, or is that just a subset of what they're doing?

Jason:        Well, I think that's partly the issue, is that I think a lot of teams in some ways see their remit as designing nudges rather than providing a broader toolkit to the table.

                There was a debate involving George Loewenstein and — I'm forgetting who he co-authored the paper with — and then Richard Thaler, and basically they were asking this question, around are these policy applications of behavioral economics simply being too narrow and focusing on nudges? Thaler comes back and goes, "Oh, well, look. This is a ridiculous complaint. In 'Nudge' the book, we never — him and Sunstein — never said that nudges were the panacea, the only application."

                But I think when it comes to practical application, quite often they are seen as the panacea, or the toolkit that they can pull on.

Julia:        I see. You mentioned that one other alternate angle that you considered for this article was the question of whether behavioral interventions that look impressive in isolation are less so if we consider the system wide effects. What did you mean by that? What would that article have said?

Jason:        Sure. Well, there's two layers to this. I think one layer people know about but haven't really looked at is, I suppose, going, "How does the intervention affect someone's behavior at a broader level?" Think about putting comparisons with people's neighbors on their power bills, to see if they're going to reduce energy usage. The experiments are generally run where they look at what happens to the power bill of those who received the comparison and those who didn't.

                But the bigger question, and really the objective, is going, "How do we reduce overall energy usage?" So, are these people, after reducing their power usage at one angle, are they then, say, either morally licensed or simply more flush with cash, that they then do other activities that increase their energy usage? That's the first layer.

I think a lot of people in behavioral economics, behavioral science, have seen that. It's just probably a bit more of an open question.

But then there's the next layer going, "Well, what's the actual effect on people's, say, mental health, their happiness, on being subject to these comparisons?" All these experiments, they'll send out these letters, and a 5, 10% shift in behavior is a real success, seeing a really small proportion of the population change behavior.

But what about those who don't change their behavior? Are they better off, or are they worser off because of those comparisons? Even in some areas such as, say, debt collection or tax collection where people may have really tough constraints and reasons why they haven't paid, does this now social comparison going, "By the way, you're a deadbeat and not paying your taxes," how does that affect their mental health?

I saw one academic at a recent presentation suggesting this is the sort of thing that could have serious mental health issues, to be constantly comparing people to others when they may not necessarily have the ability to respond and change their position.

Julia:     So, that's kind of a critique of nudging still within a utilitarian framework, of what are the consequences for people's well-being?

Jason:     Indeed.

Julia:     There's a different kind of critique that steps outside of the utilitarian framework and says, "Do we have a right to nudge people in this way?" Even if it's very likely that it is going to make them better off, are we depriving them of some autonomy?

I think a lot of laypeople, not just ethical philosophers, feel a bit of discomfort around the idea of nudges as being in this gray area, of manipulating them subliminally or something like that. Or, forcing their hand in a way that's not visible to them.

Do you have any sympathy for that concern?

Jason:     I do, at a general level. I suppose there's a lot to be said for the idea of autonomy, where people, they're deliberating and coming to decisions on their own power, rather than being manipulated towards those. Despite the fact that they have freedom to move away. At a general level, it does concern me.

I think what's kind of interesting is when you get down to each individual nudge, and Cass Sunstein makes this argument in response to some of these claims — when you go nudge by nudge, a lot of them, they don't seem quite as problematic as that overarching angle would appear.

I do see why people feel uncomfortable. I feel it. At the same time, a lot of nudges, at the end, I'll go, "Okay, that's probably not that big a problem."

There's various reasons for that. Sometimes you can't avoid having a frame, or an order of presentation, whatever it may be. You're always going to be subject to some form of influence.

I think it's things like that, we're going "Okay, it's probably not so bad that people choose a non-random ordering. It's not truly as coercive, or autonomy-destroying, as perhaps the idea of trying to deliberately manipulate people towards something, when in fact they would have had a more neutral frame beforehand."

Julia:     I mean, are there any kinds of nudges that you would have qualms about?

Jason:     I suppose the ones which I have the biggest qualms about, is probably around the use of a lot of the opt-in, opt-out choices. They're probably one of the most powerful ones of nudge.

In fact, I suppose when you go, "Well wait, what's something that's going to really stand out, to the test of time?" it's the idea that the people don't tend to change a choice, the designated opted-in or opted-out choice.

There, of course, you really do have this case where, they may be placed in to an option which, if they were engaged and asked to rationally think about it, they wouldn't do so. Think about organ donation. They have this system in a lot of Europe where you're opted in as an organ donor. There, by opt-in, it particularly means: If you're a citizen of that country, and you haven't gone and lodged a form at the relevant department, you are considered an organ donor.

That, to me, is fairly pernicious in some ways. It's having people labeled, when they've probably never given thought to it. Richard Thaler, even he has said that as far as he's concerned there's a far better option here — that's active choice. People are made to make a choice at certain points of time.

Julia:     I mean, that seems like even more autonomy than the default, which was opt in. People weren't even aware that there's a choice there. This seems, like, just strictly better. Increase people's autonomy — and also, probabilistically, increase the number of organ donors. So why not do that?

Jason:     Indeed, indeed. Also, coming back to my comments on social norms — I think, the use of social norms is also an interesting one. There is this question of, "Are there simply negative effects on some people, that we're not currently measuring?" I think there's some interesting questions here, around effectively, he's trying to manipulate people, not for their own good. Even though, they might have a bit of benefit.

But, it really is a case of "the government wants a certain result, and they're trying to shift people one way, or the other"? Once that line between, "is this for someone, for their own good, as judged by themselves?" … That's, I guess, the mark that Thaler and Sunstein set down, for a good nudge. In application, is this really for the good of the government? Are they really the ones who are the interested party here, and the ones trying to get the result?

Julia:      Right. I think, in one of your most recent blog posts, you referred to someone else's theoretical critique of nudging, I think it was Sugden, I might be mispronouncing the name.

His critique, which I didn't fully understand, maybe you can explain better, was around the idea that nudging assumes these latent preferences. That people have these true preferences, which the nudges are helping them realize. And which they otherwise would not be able to realize, because they have this shell of irrationality around those latent true preferences.

He argued that's not really a good way to model people's psychologies. Can you explain it better than I can?

Jason:      Yeah. In some ways, his complaint comes back to this idea we talked about a bit earlier, around how economists and behavioral economists are just really reluctant to let go of this rational, expected utility model of humans. The way that most of the approaches and the behavioral economics for that, theoretically go, "Okay, what can we do?"

They end up in this world, where they effectively assume that inside we actually have this rational person, who has the full set of preferences against all their choices. It's almost a psychological little shell around it that has lack of attention, lack of computational power, lack of will power, that leads to those inner preferences not being realized. A lot of the nudging is almost framed as an attempt to allow people to realize those inner preferences.

Bob Sugden's critique of this is really around the idea that basically these latent preferences don't exist. There is no internal, rational agent that's coming up with them.

He uses an example of, let's suppose that you've got someone and they're going to a cafeteria. This, again, drawing on Thaler and Sunstein's work, where the ordering of the food can affect their choices. You put more cake at the front, and they'll buy more cake. I'll put the salad at the front, and they'll buy more salad.

Sugden says, "Let's imagine you've got someone going in there, they actually don't have a preference between cake and salad. They're largely indifferent. But they simply feel this urge to eat it regardless. That's this person. Let's imagine another person just like them, but they have none of these constraints. They have no will power constraints, they have no constraints on their self control. They have full computational power.

They can calculate anything. Let's suppose that same person goes in but has the same preferences. If they don't have this initial instinct of, "I want cake," or, "I want salad," then, what's actually going to happen is when they walk in to that café, they're only going to have the same feeling as that original person. It's no failing of their rationality that's leading them to go, "I'm going to eat the cake, that's the first." It's simply the fact that they don't have this well formed preference before hand.

This seems a little bit nippy, that example. I got to the end of it, myself, and I went, "Is this really important?" In the end, do I care? Is this a bigger problem?

The more I think about it, I guess, the more sympathetic I am to it. Although it's a trivial example, it points to this fact that we probably just don't have really well formed preferences internally. There is no psychological mechanism by which they're formed.

It's clearly this "as if" model, just trying to create this idea that we make decisions *as if* we're a rational person inside this psychological shell. That's clearly not the case. There's probably a lot more serious decisions, where we're not going to have this nice, stable, internal, latent preference.

There's another paper which he talks about this idea of someone's going in to get surgery. This is again, debating Sunstein ... Bob Sugden goes, imagine this person. He's in there, and he gets presented the surgery options, and it's in a "gain" frame. They say it's 95% chance for survival. I suppose it's the right choice. They're not going to say 5% chance of death. They give him the 95% chance. He goes, "Yes, I'm happy. I'm willing to have the surgery."

Then what happens? Later on that day, he then gets shown the loss frame [of 5% chance of death]. He gets shown a different frame and suddenly, he's like, "I'm not happy about this anymore."

Which of those is actually the true preference? I think it's a case of, there's not some true preference you've got to tease out of this person. Literally every time they exhibit this preference, or come to it, they probably are influenced by the framing of that choice, so to say whether this person wants to do the surgery or not. There's no truly rational choice inside that person.

You'll never be able to really pin it down. It's like, we're going to have to either go, "Look, we're not going to do it," as judged by themselves, by this inner rational person, or we're going to have to find some other way of justifying the nudge that we apply.

Julia:      I see. If we can't justify it by saying it's just realizing their own true, deep down preference, then we would have to justify it by, "We're maximizing their well being." Or, something like that.

Jason:      Yeah. Effectively, we'll have to take a somewhat more paternalistic reasoning behind what we do. It's essentially saying we believe this is in their best interest, based on the fact that most people want to maximize their chance of survival, or whatever it might be.

Julia:      I wonder if there was some kind of thing you could put on your official drivers' license, or whatever, government ID, that's like, "I consent to being subject to these kinds of nudges," or something. Then, governments and companies would be allowed to nudge you, since you opted in, in general, to nudging.

Jason:      I agree. I think, for me, one of the easiest ways to overcome a lot of these problems, is simply through the idea of consent.

Julia:      I just thought the problem might be that if you ask for consent in any individual case, it might ruin the effect of the nudge, or something. This would just be a broad consent.

Jason:      Yep. Possibly. Although, there is a lot of evidence that telling people about nudges doesn't actually diminish their effect as much as you might think. There is a bit of research on that point, which is hopeful.

You're probably right though. Perhaps, some of this isn't, they're going to say, "I'm going to consent to this at a fairly broad level." ... Around technology now, Apple could say to you, "Okay, do you want to sign up? We're going to nudge you towards purchasing our products or better music." Or, whatever it might be. "Do you agree to us trying to work out what your better interests are, and guiding you towards this, or would you rather just have us serve you something where we haven't tried to make that judgment?"

Julia:      Right. Although, I guess they could always make that question opt out instead of opt in, and then it would be nudging you towards accepting their nudges.

Jason:      Indeed, definite regressive.

Julia:      In a recent episode with Chris Auld, we had a parallel conversation. We were discussing critiques of economics in general, as opposed to behavioral economics, and behavioral science. In that episode we started out complaining about bad critiques of economics, and then later transitioned into, are there any good critiques?

In this episode, we've been talking about behavioral science, and we've been talking mostly about good critiques of behavioral science, critiques that you hold or sympathize with. Like: "It's so trendy, behavioral insights teams have neglected standard economic approaches in favor of these sexy and easy behavioral insights, nudging approaches." And "Behavioral economics has been cataloging these biases without attempting to explain them." Things like that.

Are there are any, in your opinion, common and bad critiques of behavioral economics? Areas where it's been straw-manned, or misunderstood?

Jason:      Yeah, I think a lot of the original critiques actually coming out of economics itself, are probably not particularly good. I think some try to extend the defense of the rational economics too far.

There's this book by David Levine, "Is Behavioral Economics Doomed?" — and on the one hand I take a lot out of that book. Again, relating to my earlier points, there is actually a lot of power in that rational actor model at times.

But there's so many points where you have to concede defeat. Or, really say "There's something going on here that isn't matched." The example of costless opt in, opt out, is a great example where — surely at that point, you're going to say there's proof in the pudding, in the changes in retirement saving through the Save More Tomorrow Plan. Just concede that there's something there. But they don't want to let go.

Julia:      They won't concede that there's something that can't possibly be consistent with a rational actor model, you're saying?

Jason:      Indeed. Even Greg Gigerenzer goes there at times. Back to that 1994 set of papers, debating with Kahneman and Tversky. In the back and forth, he goes, "Look, it can be explained it this way." Then, Kahneman and Tversky come back and go, "Actually, look. We looked at that a few times. It didn't quite work out that way."

            But he just didn't want to let it go. I think ultimately, you have to accept that there's clearly some deviations. There's clearly some problems.

            The question though, the interesting point of the debate, is where exactly should the line be drawn, and how exactly should we think about those problems. That's a far more fruitful area to discuss.

Julia:      Yeah. That's probably a good place to wrap up — Jason, as you know, if you're a fan of my podcast, I like to ask my guests at the end of the episode to nominate a resource, like a book, or a blog, or article, that they have some substantial disagreement with, but they nevertheless recommend — in the sense of, "this is worth engaging with, and thinking about." What would you nominate?

Jason:      I'm going to nominate Angela Duckworth's work on Grit. There's a broader group, which I've really been engaging on this point. There's some other people I could have named. James Heckman, or Caroll DeWitt.

            But really it's around this idea of: To what extent are there these other traits — in Duckworth's case, it's around grit — which affect success? So, how much do they really affect success? Then, in turn, how malleable are they?

            I'll probably have a couple of points, but I really grapple with her work around the question of how important is grit, really, towards success? Is it really just conscientiousness repackaged? Is there something genuine here? Similarly, to what extent can we change someone's grit through training?

            I suppose, looking through a lot of literature, I struggle to see the power of these interventions.

            But, at the same time I find the way she engages with her critics, is really productive. I quite like it, when she responds to the critics in news articles — or in fact when she was pressed... There was an episode of EconTalk where she was interviewed on this. I think she's a real model of how to deal with critique.

            But just her work in general challenges a lot of my pre-existing biases, partly coming from the evolutionary angle of some of my previous thinking, of how malleable are people? Can we really change them and their outcomes in a substantial way? Particularly through, I guess, smallish and costless interventions — can they really make a big difference?

Julia:      Right. Interesting. That's the perfect recommendation for me, in particular. I get so excited to find people who are good at responding to criticisms of their work, in a fair and nuanced way. That's a red flag — in the sense of waving a red flag in front of a bull — for me, to check out her responses to critics.

I guess we can link to that episode of Econ Talk with her, and if you have any other suggestions for examples of Angela Duckworth's handling criticism, or disagreement, you can pass them on to me. We'll link to those on the podcast site, as well, and maybe to just one of her books on Grit.

Jason: Indeed.

Julia: All right. Cool. We'll link to those, and we'll link to your excellent blog, and several of the papers that came up in conversation, like the 1994 Kahneman-Gigerenzer debate. I think that's it.

Jason, thank you so much for coming on the show. This has been an interesting conversation.

Jason: Yeah, it's been a real pleasure. Thanks.

Julia: This concludes another episode of Rationally Speaking. Join us next time for more explorations on the borderlands between reason and nonsense.