

RS141: Dan Sperber on “The Argumentative Theory of Reason”

Julia Galef: Welcome to *Rationally Speaking*, the podcast where we explore the borderlands between reason and nonsense.

I'm your host, Julia Galef. And with me today is our guest, Professor Dan Sperber. Dan is a French social and cognitive scientist, who specializes in the fields of anthropology, linguistics, philosophy, and psychology. Currently, he is a professor of philosophy and cognitive science at the Central European University in Budapest. Dan, welcome to the show.

Dan Sperber: Thank you for having me.

Julia Galef: Today what I want to talk about is: there is a standard view about the human capacity for reason, which is that it evolved in order to help humans figure out the truth and make better decisions. Which in theory should produce better life outcomes, and is therefore adaptive.

However, there are a bunch of puzzles surrounding this theory, which cognitive scientists have increasingly noticed and documented over the decades. In that our capacity for reasoning doesn't always go as it should.

Dan Sperber, along with his colleague Hugo Mercier, have developed an alternate theory about the human capacity to reason, called the argumentative theory of reasoning. That's what we're going to delve into today.

Dan, would you like to kick things off -- by either stating your theory and then elaborating a bit, or maybe giving some of the context for what motivated the development of this theory?

Dan Sperber: Well, yes. Let me first point out that the idea that human reasoning is for reaching better beliefs and better decisions predates any idea about evolution. The idea of humans as a rational animal dates back at least to ancient Greek philosophy, and it's a cliché that you find again and again that what makes humans superior to all animals is precisely this capacity they have to reason on their own, and to reach better decisions and form better justified beliefs.

People have had very great ideas about that. Maybe the most extreme one is Rene Descartes, who thought you should forget everything you learned from others. And rebuilt his knowledge of the world just by the use of reason, accepting as true only things that he could justify with reason. This is such a grand project -- grand to the point of absurdity,

Need Help? <mailto:support@rev.com>

especially if one thinks of, indeed, the limits of reasons, of reason, and reasoning. That, as you pointed out, have been very well documented by now in psychological and cognitive science and research.

When evolutionary ideas started developing with Darwin, basically they just took up this whole idea, that the evolved function of reason is to go beyond the limits of perception and automatic inference that are found in other animals, to go much further in our cognitive achievements. You can find quotations in Darwin stating that. And in much of recent literature in psychology or philosophy, again this is just taken for granted. That the function of reasoning is basically for individual cognition, of individuals pursuing their own goals.

What we argued, Hugo Mercier and myself, is that this is a very implausible story. And a story that doesn't square well with the evidence we have.

Before even looking at the all the experiments that have shown how poorly humans often reason -- how badly they perform on a great variety of apparently simple reasoning tasks -- there's a kind of more basic and simple enigma.

Take most of our cognitive capacities. Take vision, for instance. If you're not visually impaired, you take for granted that other people who are not visually impaired either see the same thing you see. If there's a tree in front of you, you see the tree, they see the tree. If there's a dog, you see the dog, they see the dog. If there's a sunset, same, you see the sunset, they see the sunset.

The notion that our perceptual abilities -- not just at the margins, not when we exert them in difficult circumstances where the light is not very good, or we're looking at a very distant object, but just in normal circumstances -- could cause us to have the divergent perception, is very weird. Yet, this is clearly what happens with reasoning.

When people reason, say, about political issues or moral issues -- not only does it *not* cause them to converge, but often (and there's a lot of evidence to that effect) it causes them to accentuate, exaggerate their disagreements.

If reason was a way to reach true knowledge, it should, to begin with, cause us to converge, to come to the same conclusions. Again, if you ask people to look at something and say what it is, you expect them to, if they know the object that it is, to come away with the same judgment. If

you ask people to reason on a problem, quite often you don't expect any such conversion. That's a big enigma. It's not compatible with the idea that reasoning aims, as its function, to present us with deeper and better knowledge.

Then, indeed, there's all this experimental literature that shows that many simple, logical tasks or reasoning tasks -- that once you explain them seem totally obvious -- in fact cause failure, again and again. That's quite puzzling too.

Now, if we enter into the nitty-gritty of psychological research, one can start discussing each experiment and say, "Oh yes, well, people are not really making the mistake you think they're making. They interpret the task differently," and so on. We can defend the performance of people participating in psychological experiments and say that they are being tricked by the psychologist, trying to cause them to fail by presenting kinds of cognitive illusions. I think that defense of the ordinary participant in experiments is reasonably fair.

Still, psychologists might also say, "Of course, we are provoking participants to mis-reason, to reason in an inappropriate way, but that's the same thing as using visual illusion in order to study vision. It's precisely when we explore the limits of the system that we also understand its basic functioning." That works for some kind of problems in the study of reasoning, some of the famous experiments that were devised by Kahneman, Tversky, and others and so on.

Julia Galef: To give an example of such a case... you can tell me if this is a good example, but I'm thinking of things like: tests of how people use probability, showing that people reason poorly about probability.

Defenders of human reasoning will say, "Well, you're using percentages, which are not a natural way for people to think. When you rephrase your problems in term of frequencies, relative frequencies, people do much better."

Dan Sperber: Yes, that's a good example, and there's been a big debate. There's been answers to that too. So that was the German psychologist Gerd Gigerenzer -- a very interesting reply to Kahneman and Tversky.

There are even simpler cases which look like reasoning mistakes. To take one of the simplest, well-established type of results, if you ask people to reason from a conditional statement, like ... "If there is a square, then it's blue," then you say, "There is a square, what can you infer from it from

this premise, 'If a square, then it's blue; There's a square,'" Then they infer correctly that it's blue.

Now you tell them, "If the form is a square, then it's blue (I should rephrase it like this)," and that the second form is red. What can you infer from that?" Then only half the people correctly infer that if it's red, it's not a square.

If you tell them it's not a square, half the people infer it's not blue, that's a mistake. If you tell them, "It's blue," they infer it's a square, which is again a mistake. Very simple logical thoughts like that seem to trick people in a dramatic manner.

Again, that's not what strikes me as the most dramatic type of error. Because there are ways to discuss what people are actually doing, or thinking they're doing, while they answer the way they do. Basically, I think what happens in such cases is that people do not accept the instructions of the experiment, which is to reason in a strictly logical manner. Because that's not what we do in life, in the strictly logical, in the sense of deductive, manner.

When you reason in your daily life you take into account whatever is relevant. Whether it has deductive consequences or not is beside the point. We use both deduction and non-deductive reasoning, probabilities, we mix it all up together. What we're not good at is separating, isolating, just the logical deductive aspects and ignoring all the rest of the information. That needn't be a failure of reasoning. It may just be a failure of performing a certain kind of logical task that you never have to do in an isolated manner in real life.

Take another type of failure, which is much more striking in a way, and puzzling, and not that easy to explain. It's a series of failures which are often described as being a confirmation bias. What people do... if you give them a problem to solve and if they have already some hypothesis about what is the solution of the problem, or if they just get a hunch just when you present the problem -- then what they end up doing is piling up evidence and reasons that confirm their belief or their hunch, and ignore or not pay attention to all the evidence or arguments that would go in the other direction.

There are lots of experiments where people have an initial hunch on some problem. Like the famous Wason four-card selection task, I'm sure many of your hearers will have encountered it before. People get a hunch and then you give them the time to reason, think harder about it. And

they spend the time you give them, not to really go over the possibilities, see whether their hunch is really justified, but just developing the best possible argument to justify their hunch. Even when their hunch is mistaken, and therefore the best possible argument is pretty lousy.

We get this confirmation bias, and that affects not just experimental settings, but real life. We all have the experience of discussing with friends, and being really aggravated by the way they just stick to their guns and keep not seeing our excellent arguments, and just piling up poor arguments for their point. And of course they think the same thing about us. God knows who's right.

What happens, quite systematically, is this tendency to use reasoning not to discover the truth, not to examine in an open-minded manner all the possibilities -- but basically to find arguments in favor of views which are held even before reasoning. You see this in our social media environment, or again, because we just have a kind of immediate intuition that we hardly depart from once we have it. It's as if reasoning played the role of a lawyer whose job is to defend your views, whatever they are. And your actions, whatever you did. Not to discuss if they were the right views or the right actions.

Julia Galef: Rationalizing, then, is not a bug in the process of reasoning, rationalizing is the purpose. It's the feature of reasoning.

Dan Sperber: That's a direction in which the argumentative theory goes.

Again, the problem is very well known, it's quite central to work on reasoning. It's typically most commonly called the confirmation bias. But we prefer, as also modern psychologists do too, to call it the myside bias. Because... what if you ask people to reason about some view that they disagree with? Then they are quite good at finding counterarguments.

It's not that if you reason from a certain view that you automatically have to confirm it. If you adopt that view, if it's your view -- and again, it doesn't have to be a deeply held view, it can be just a hunch -- then you will indeed go for this type of confirmation. If it's a view you disagree with or you have a hunch against, then suddenly you find yourself very, very bad at confirming, and very good at finding counterarguments. It's a myside bias.

This raises a huge problem for any view of reasoning as aimed at improving individual cognition. Because that's exactly what you don't want to do. You want to have an open mind on issues, even be ready to

revise your own views. To weigh evidence and arguments in the most objective kind of manner. Objectivity is in itself a condition for getting to better knowledge and better decisions, and that kind of objectivity people don't have. We all think that we have it, but if you look at the evidence, we don't. It's puzzling, truly puzzling.

What's been the reaction of psychologists and philosophers to the puzzle has been to either say, "Oh well, it's not as bad as it sounds and there are cases where in fact it's a good reasoning strategy." Which I agree is true, it can be good to try to defend a certain view you have if it has a certain kind of initial improbability.

But again, all this myside bias is much more pervasive. It's certainly not limited to a few cases where it would be useful. People have been happy to say, "Oh, you are not very good at reasoning. Ha ha ha," as if it was some kind of newsworthy finding... they just grab on our own mediocrity and stop there. You find a lot of literature, in popular books and so on, about reasoning, that just cite one experiment that shows that people are reasoning badly, and with lots of exclamation marks, and as if this was a great finding.

No... you can't stop there. Certainly not, for instance, if you take an evolutionary view about our cognitive capacities. If we reason so badly, then the people who don't, who just stop reasoning, would do better and have more descendants. And we would be the descendants of people who have stopped reasoning.

Julia Galef: What about the argument that there's some degree of optimal reasoning, or some degree of reasoning that's optimal from an evolutionary standpoint, but that's not 100%, because reasoning is costly.

Dan Sperber: The argument is that reasoning is not perfect from a strictly normative point of view because there's a cost to it, a biological cost, and of course we settle for the optimal cost-benefit balance rather than aiming for the perfect reasoning from a logician's point of view. That is a perfectly reasonable argument.

But it doesn't explain something like myside bias, or the benefits of being biased in this manner, or the costs we would incur if we were not so biased, have never been demonstrated. In fact they're very hard to imagine what they could be. It's not particularly costly in itself to be objective, and it's very costly not to be. For example, in perception we're objective, we don't see what we would like to see, we see what there is.

Julia Galef: In visual perception.

Dan Sperber: At the visual level. When we interpret it we may be biased. In a lot of basic cognitive processes we're not biased in that manner. Animals are not biased in thinking that there is food where they would like it to be. They're biased to think food is where it's likely to be. That sort of explanation.

The kind of explanation that the argumentative theory that Hugo Mercier and I have tried to develop proposes, moves in a radically different direction. Saying, reasoning didn't evolve for the benefits of the individual. It's not a means to improve our individual cognitive capacities. It has to do with the fact that human beings depend so massively on communication and the role it plays is crucially linked to communication.

How does that go? It goes in this way: indeed, human beings owe so much of information they possess, and use every day, to communication with others. You cannot become a competent adult without having communicated a lot with your parents, caretakers, peers, and so on, and you cannot spend a day without drawing on information that you got from others. We stand to benefit from communication to others much, much more, incomparably more than any other species.

Our communicative capacities are quite extraordinary. At the same time they make us vulnerable to misinformation, to manipulation, to deception. And it makes it also advantageous for us, quite often, to mislead others, to deceive them. To misrepresent things, to protect ourselves, to influence them in whatever way we want.

An argument that a number of my collaborators and myself have made is that this has created a selective pressure. Not just one mechanism, but for a range of mechanisms of what we call epistemic vigilance. Mechanisms that aim at trusting information when it indeed it is reliable as much as possible. And sifting the information which is presented so as to accept that which is most likely to be correct and reject that which is not. Without such an optimization, the tendency to optimize the receptivity to information received socially, the risk of being victims of communication rather than beneficiaries of it would be great.

This epistemic vigilance takes two basic forms. There's the issue of who to trust and who to believe, and the issue of what to believe. First, we are very good at judging to what extent people are trustworthy on a given matter. When they're competent, when they're benevolent. And studies starting from early childhood show that this is a very systematic feature

of our social interaction. We care very much about the trustworthiness of others, and we're not too bad at judging it.

The other issue is what to believe. Some things, even from somebody who is trustworthy, are more believable than others. If I tell you now that the world is about to end in the coming five minutes -- even though I'm a very trustworthy person, I hope you trust me in every respect -- even though I'm such a trustworthy person, you wouldn't believe it, and so on. Some contents are more believable than others.

One in particular way in which we judge the contents of trustworthy is we see: are they internally consistent, don't contain any kind of internal contradiction? Because they shouldn't. And second, are they consistent with things we already have strong reasons to believe? We look for the consistency of new information with what we already believe -- having sometimes to revise our own beliefs, but not very easily and often more likely to reject new information that contradicts what we believe.

There's a kind of subtle capacity here we have to take in new information, sometimes at the cost of revisions, but also sometimes rejecting information that is inconsistent with things we have strong reasons to believe. If we are consistency checkers, this means that we can resist a lot of misinformation, deception, lying. Because typically false information is inconsistent with a lot of true information, and that comes out.

Now, I am trying to convince you of a theory, the argumentative theory of reasoning. You and the audience of your podcast, you might say, "Why should I believe it?" Certainly, you shouldn't just believe it because I'm a nice Frenchman, that's not good enough. Maybe it should be, but it isn't.

The way in which I might convince you is by showing you that it would be more consistent for you to believe it than to disbelieve it. To show you that given the number of things that you already believe, given further experimental evidence and so on that I can put forward, the consideration where I am trustworthy, I'm a psychologist, I'm a philosopher, so I can tell you about experiments and you'll believe me on that quite nicely...

Given all this, it becomes more consistent for you to accept a new theory than reject it. The idea is that reasoning that is displaying the consistency of relationships between a set of premises and a conclusion -- showing that if you accept some premises, you accept some conclusions -- is a way to overcome the epistemic vigilance of others. When we talk about

matters where we don't have sufficient authority to get what we say accepted just on the basis of our authority. Reasoning is basically, from that point of view, from the point of view of the communicator, aimed at succeeding at communicating information that would not be accepted just on trust.

From the point of view of the audience, we find ourselves in a situation of, on the one hand knowing that there may be good information for us that people could share with us, but finding that they are not trustworthy enough on these issues to accept what they say just on that. We're missing out on some good information because it's too important to accept it just on trust, and still it could be true. If others can argue for the information they want to share with us, ideas they want to share with us, goals they want to share with us, show us that if we aim at having a consistent set of beliefs and rational decisions then we should, in fact, be convinced by what they say.

It's their benefit and our benefit, so reasoning is used in two different ways. On one hand, by communicators who argue for ideas that they want to share, that's the argumentative part of reasoning. From the point of view of the audience, the nice thing about arguments presented by others is you can evaluate them. So the audience is in an evaluative position.

One prediction, for instance, that's in a nutshell the basic theory we're presenting, the argumentative theory -- is that reasoning is a means to overcome the limits of trust in human communication. By giving reason for others to believe what we say, if we're on the communicator's side; or by evaluating the reasons others give us, in order to be convinced on good reason and not otherwise.

Julia Galef: So the formation of the belief within the communicator is not produced by reason. But the propagation of that belief to other people, hopefully successfully, that's the function served by reasoning. It's figuring out how to transfer your belief to other people?

Dan Sperber: Yes, it's filtered by reason. The way our acceptance of ideas and information from others is filtered by reasoning. Again, from the point of view of the audience, we could have called it the evaluative theory of reasoning. From the point of view of the communicator, it's the argumentative theory of reasoning...

One prediction, one very actually surprising prediction of the argumentative theory, empirical prediction that can be experimentally

tested, is that this myside bias, this confirmation bias, is going to be found quite systematically in the production of argument, but not in the evaluation. People are going to be much more objective when they evaluate arguments than when they produce them.

Julia Galef: But aren't there many cases of myside bias demonstrated in the way people evaluate arguments? That people are more likely to reject arguments -- even exactly the same structure of argument -- if it supports a position that they hold they'll accept it, and if it supports a position they don't hold then they won't accept it, basically?

Dan Sperber: That's indeed what would have been the common wisdom a few years ago, and it really seemed to be the case. But if you look more closely, that isn't so. It depends on which kind of situation you find yourself in.

If we are in the situation or in the conversation where I have some new ideas that you're willing to pay attention to, you think it may be interesting but you're not willing to believe them on trust, then you would evaluate objectively. And if I argue well, you would be convinced, and otherwise not.

If we are in a debate, like a political debate or a moral debate, where we start from quite antagonistic views that we strongly hold, and we don't have kind of a shared goal of indeed finding the best solution to a problem or finding what the best theory is on a certain issue -- in that case, indeed argumentation tends to provide polarization.

In that situation the audience is not in a situation where the problem is, "Is there some useful information or ideas that I could benefit from?" The position, the attitude, is "How can I fight against the views from the other side?" Indeed, in that kind of situation it's polarizing both sides. It's myside bias both in evaluation and production.

I'm talking about what I think is a basic situation which drives the evolution of reasoning and argumentation where there's, in fact, a joint interest of a communicator and the audience in converging on an understanding of good ideas. What the person who argues in reason does is say, "Okay, you're not going to believe me just on the basis of my authority, not just on trust -- but I'll show you that in your *own* thinking, if you look at my arguments, you'll find that they are good and you'll accept them."

Of course, there's a lot of bad arguing... like sophistry and so on, the process can be misused, but it does basically allow the transmission of a lot of information that wouldn't be accepted otherwise.

The way we developed this idea -- first we revised and reconsidered, revisited, a lot of past evidence. And shown that to the extent that the difference between objectivity, and production and evaluation of argument, has been studied at all, there's plenty of evidence already there showing that people are much better at evaluating than they are at producing. Now, with our theory we've developed is creating new research aimed at testing that directly, and indeed finding pretty good confirming evidence.

To give just a very simple type of finding which is quite decisive: if you take logical problems. I mentioned before the Wason Selection Task, which is a logical task, a simple one, but where only 10% of people normally succeed. If instead of giving it to one individual, or to individuals separately, you give it to groups of three or four people, and allow them to solve it together.... It's an experiment which has been replicated many times. What you find is that about 80% of the groups succeed, when only 10% of individuals succeed. What happens is that people come with different views. Often none of them has the right solution. But each individual presents his views, or her views, which are being evaluated more objectively by the others, who will reject the views which are not really sound, not really consistent, and so on.

All the values get sifted out and people progressively converge in conversation, in discussion, on the right solution. That's a situation where each group is trying to see. There's a common interest in finding the right solution, and they do very well.

This kind of experiment converges with what's been found, for instance, in teaching... When you ask pupils, students, to resolve a problem in science, when some will want to try to understand, develop the ideas on their own -- they do, they learn better, they come to the right solution on their own better than if they're just told the solution to a science problem, say by the teacher.

In political interaction you have what's now called deliberative democracy. If you get a group of people and give them the task of solving a shared problem together, they may approach it from very different views. But if they have to interact and listen to each other's argument, again, they come with very biased arguments but they [are less likely to] accept the biased arguments of others. So progressively, provided that

they have this common goal, they converge on, possibly not the best, but on a good solution.

Julia Galef: Dan, where is the selection effect that you're positing operating? Does the argumentative, or evaluative, theory of reasoning persist because it's useful to the individual? Because it's useful to the individual to convince other people in his tribe? Or does it persist because it's good for the group, for people to be epistemically vigilant but also good at arguing?

Dan Sperber: I think it's very much a matter of individual selection. The benefit is for each one individual. The benefit is for the communicator who has an interest in influencing, and convincing others, but has to pay the price of producing good arguments for that. And the benefit is for the individual recipient who has an interest in getting good ideas, sound information...

From a group point of view, the issue is it's not clear that it has a group interest. In that, first, because if you assume that the transmission of information in a group has to do with group selection, has to do with the benefits of a group in general, then trust should be dominant. There shouldn't be the problem that I've mentioned in terms of the limits of trust. There should be no need for epistemic vigilance, or a much lesser need. The very need for reasoning, the very need for argument, speaks to the limits of our mutual trust.

To go to the extreme, if you think of a society of insects, like bees, who are transmitting information from one to another, they don't need to argue. They trust each other completely. But they are biologically completely aligned, they are clones or near clones. Their interests are identical.

No, the argumentative theory is a theory about evolution of individual advantage. Which, in itself doesn't say anything about the existence or non-existence of group selection. A number of colleagues think that group selection has been quite crucial to the evolution of humans; they typically argue for a multi-level selection, nobody believes it's only group selection. The point about the argumentative theory is a theory about processes at the level of individual selection.

I've been speaking about the positive effect that group discussions may have. They have a positive effect only when people have a common goal of reaching a good decision or finding out what the truth is. Otherwise, as we mentioned before, you can get argumentation causing polarization, people coming with divergent views and ending even with more

divergent views... Again, there's a lot of social psychology evidence to that effect.

Now, it's not that reasoning is good for the group. Sometimes it's good for some groups who have a common goal, but not for other groups in other kinds of conditions. What it's systematically good for is the individual in the way they try to benefit from communication from others, and in the way we try to use that communication to benefit from a relationship with other people.

Julia Galef: Okay, can we go back briefly to the epistemic vigilance concept?

Dan Sperber: Yeah.

Julia Galef: The idea is that we evolved the capacity to reason so that we could get through other people's epistemically vigilant defenses. A lot of the listeners of this podcast are in the skeptic movement or the skeptic community. And one of the big frustrations of skeptics is that people seem to very easily believe things that they read on random websites of the internet, or that they're told by people with no credentials whatsoever. Skeptics are quite used to bemoaning that people are not epistemically vigilant. And I'm wondering if there's a tension there with this idea that people automatically check things that they hear against their current beliefs, because it doesn't seem to match the behavior that I'm pointing at.

Dan Sperber: I'm a skeptic too, and I bemoan too the fact that people are not epistemically vigilant enough. Typically these are issues of opinions about extraterrestrial beings, unidentified flying objects, theories about the assassination of Kennedy, and so on, and so on. A number of issues which can be very entertaining and feed conversation endlessly.

But the same people who are gullible on such issues, if you try, may be just as vigilant as the next person about the child who tries to pretend that she has done her homework when she hasn't, or a travelling salesman trying to sell.

The social benefits of all these public beliefs, I think, which have little bad consequences at the individual level, may have very bad consequences at the social level. Things like global warming, for instance. The denial of it is going to cost us all.

In terms of individual choices, we don't drive people to be particularly vigilant. They find a greater reward and benefit in holding what they

think is a daring position, showing that they don't fall for any of these ideas or authority of the scientist, or whatever, whatever. That's important and highly regrettable and so on, but it doesn't directly provide us evidence against the view that it's a basic aspect of human interaction that we are vigilant, that we do not accept everything others say. Even people who are highly gullible in some issues are not that gullible in their daily life. Total gullibility is a pathetic pathology that is very rarely found.

Julia Galef:

Would you say that the argumentative theory is in tension with Kahneman and Tversky's work? As explained, for example in *Thinking Fast and Slow*? In which our system one, our intuitive thinking system uses all of these quick and dirty heuristics, that sometimes systematically go wrong in certain contexts. And our system two, our slow, careful, deliberate, reasoning system can, if we engage it, can help us notice and correct those biases? Is that compatible with the argumentative theory of reason?

Dan Sperber:

We had a conversation with Danny Kahneman after his book had come out: where will you go now? He came to a point to which most other people who have this kind of dual process view, like himself, have come to. The conclusion that in fact the so-called reasoning side, the system two type of thinking and so on, the slow thinking, is also prone to mistakes. And that the fast thinking, in fact, does very well in lots of cases.

The initial view of about 15 to 20 years ago of dual system theory, according to which you had system one heuristics, quick and dirty as you said -- it's cheap, and it works but it makes a lot of mistakes, and then, this can be improved or corrected by system two slow thinking -- That view has lost a lot of steam. I think it's wrong... at individual level.

The achievement of science -- of course it involved a lot of reasoning, but science is a collective activity. It's involved a lot of argumentation, counter-argumentation, trying to persuade others, paying attention to others because they're right and you're wrong, you better change views rapidly.

You'd have to ask Kahneman how much he agreed with our view. I think there's some incompatibility with the way, indeed, he presented the view in *Thinking Fast and Slow*, and the way the dual system theories have been developed. Allowing more technical issues that this is not the right place to discuss, yet there are differences.

Let me put that in a different perspective. There's been a lot of work on reasoning done by very bright people in the past 50 years or so, Kahneman being one of the brightest obviously. It's impressive how much has been done and how bright some of the people who have been doing it were.

On the other hand, if you compare it to what's been happening in other domains of cognitive science, what's been happening in our understanding of vision, or the psychology of perception, what's been happening in developmental psychology... You can look at so many different fields of psychology where we find ourselves in a radically different perspective in knowing and understanding much more than we did before.

But you look at the psychology of reasoning, with all these bright people - and the results are meager, to be kind. People haven't even converged on any kind of theory. It's been a field in a state of crisis from the beginning. It's just people holding opposite views and nobody every coming to ... You never got an agreement, "Oh, now we've moved forward, that's the new basis from which we start." There is a real problem with the psychology of reasoning. Not with individuals, they are very, very bright. But with the very way the field has been approached with the very [way] in which reasoning has been approached.

I'm not saying that the argumentative theory by itself will provide a new basis, and now we can indeed make huge and greater discoveries. But what I do believe is, in a way, psychology of reasoning has spent lots of time going up and down many blind allies. When so much of the rest of psychology was just making great strides forwards in a relatively straight manner. And that has to change.

Julia Galef:

Setting that aside for a moment, I had a few other pieces that I wanted to connect to the argumentative theory. To some extent, the argumentative theory seems quite consistent with observations, both anecdotally and experimental observations that are hard to explain under other prevailing models.

For example, the fact that people have this tendency to search for more justifications for things they already believe. That's not very explicable under the idea that reasoning is for helping you arrive at the truth. If you've already decided what you believe, then why would you need to find more reasons?

But then there are other observations that are harder for me to reconcile with the argumentative theory. Like the fact that there is this phenomenon, which is not universal by any means, but is present in at least some people sometimes, which is cognitive reflection, or reflectiveness. Some people, at least sometimes, can notice when their system one intuitive answer -- often to logic puzzles, this is how it's demonstrated in the studies -- people are given a logic puzzle that has an intuitive answer that will occur to you immediately, and then some people are able to say, "Wait, no. Let me sort of think about that more carefully," and then they'll notice their intuitive answer was wrong.

That doesn't seem like the kind of thing ... Clearly we, to some extent, evolved this capacity and it doesn't seem like the kind of reasoning that would help you win arguments exactly.

Dan Sperber: When 10% of people in all societies are, indeed, presented with one of these logical puzzles which figure mistaken intuition, they reflect and correct it. They've been trained to do so, it's a social game.

Julia Galef: Is it just training?

Dan Sperber: Just as we've been trained in school to abstract away from a lot of contextual information, we have the ability to decontextualize a problem from all the background information and beliefs you have. It's a skill. It's a hard-won skill, and a lot of schooling is about acquiring this skill. You know, the skill is there for a certain type of purpose, there's a lot of cases where it's not a particularly useful skill.

I wouldn't make too much of that. I don't think that this reveals any deep aspect of human psychology. Yes, I think the ability to overcome, on your own, the pull of the initial hunch, is more or less developed across individuals. Some people are better at it than others. And we live in the kind of culture where science has developed, mathematics, logic puzzles, and so on, where there's a strong pressure to acquire the ability to do that.

Even so, you still only have a minority of people are good at it. Even if you're doing many of these tests with science students, even mathematical students, economists, and so on, we fall prey to many of these intuitive biases. Unless precisely we've already been taught to recognize the trap.

I don't mean it's a game in general, I'm not saying there are no social uses. It's indeed good to be able to do that. But that is still very much part of the social practice.

Julia Galef: I see.

Dan Sperber: I see no reason to believe that this type of evidence, that you correctly point at, reveals a deep feature of the human psychology. As opposed to a deep feature of our social interaction, and enhanced in a certain kind of social context.

Julia Galef: I see. We have maybe a couple minutes left in this part of the podcast, and this might be a good place to ask you about implications of the argumentative theory. Assuming it's basically true, what does that imply about -- anything? The attempts, like my organization is attempting to train people, to think in a less biased way or to make better decisions? Is the fact that roughly 10% of people can avoid the pull of intuitive biases, is that evidence that training people is possible but very difficult? What does the argumentative theory have to say about that?

Dan Sperber: The [fact we can] can train 10% of people or more to avoid the pull of intuitive bias may not be so much a way to go. It's a bit like -- remember all the experiments on memory, you can train people to remember a long list of syllables. And these methods that go back to antiquity. But then it doesn't generalize to memory in general.

Same thing if you train people to be much better at these logical puzzles, doesn't mean that they'll reason better in their daily life, I doubt it very much. The evidence, at least, is not there.

Julia Galef: There is individual difference in rationality in everyday life, though, right? Stanovich, for example, has...

Dan Sperber: Of course there is, of course there is.

Julia Galef: But you think that's innate and not trained?

Dan Sperber: No, I'm not saying that training plays no role. I'm trying to say trying to taking a number of typical experimental designs or puzzles and so on, and getting people to see that their initial hunch was mistaken and overcome it, may not generalize the way we would like it to, to the way they use reason in daily life.

To go out on a positive note, the suggestion of the argumentative theory would be: the best way, when it's possible, to solve a problem... is to gather a group of people who have a common interest in finding the best possible solution. In finding out about facts or about theory. In getting them to come up with ... To argue, amongst themselves, to discuss. That's, in political terms, it's a very democratic way to proceed, and fairly basic. You don't delegate to a few elected people whom you're going to reelect or not reelect five years from now.

You basically, it's got to be a type of democratic interaction to be effective. In science, in a way, that's where... this conversation of people who are both competing, because they come with different views, but have a shared interest in getting better knowledge, proves itself very successful.

I would say it's less individual training and more on creating opportunities. An institution for the transmission of ideas in order to inform the making of decisions, that tap on all collective resources in this way.

Julia Galef:

Great, that's very helpful, thank you. I think that's a good note on which to close this part of the conversation, and we will move on now to the Rationally Speaking Pick.

[musical interlude]

Welcome back. Every episode we invite our guest to introduce the Rationally Speaking Pick of the episode -- that's a book, or website, or whatever else tickles his or her rational fancy. Dan, what's your pick for this episode?

Dan Sperber:

It's both a website and a book. The website is called cognitionandculture.net -- it's in the process of being improved, but it's a website that many friends and myself use, where we mix information, discussion, debates about issues which are the interface between our understanding of the mind and the understanding of society. That may be of interest to listeners of your podcast.

More specifically, we just had a book club with a very vivid discussion on what I think is one of the very best books I've read in a long time. The book is by an English philosopher and psychologist, Tom Scott Phillips. It's called *Speaking Our Minds*, and it's about the evolution of language -- a favorite topic of many, but a topic which I find is being mishandled most of the time. And this is a really novel and exciting approach. The book is

Speaking Our Minds, it's by Tom Scott Phillips and it's just been published by Palgrave. So I would recommend both the website cognitionandculture.net and Tom Scott Phillips' book.

Julia Galef: Great. We will link to both the website and the book on the podcast website when this episode goes up, and I also encourage people to check out the original 2011 paper by Dan Sperber and Hugo Mercier, *Why Do Humans Reason* -- which you can find, for example, at Hugo's website, with a lot of accompanying discussion.

Also, keep your eyes peeled for a book by Dan and Hugo called -- is it *The Enigma of Human Reason*?

Dan Sperber: Yes.

Julia Galef: Okay. Great. Dan, thank you so much again for joining us on the show. It's been a pleasure.

Dan Sperber: Thank you and good bye. Au revoir.

Julia Galef: Au revoir. I encourage our listeners to also check out the full transcripts which we post for every episode at rationallyspeakingpodcast.org. This concludes another episode of *Rationally Speaking*. Join us next time for more explorations on the borderlands between reason and nonsense.