

Rationally Speaking #206: Kal Turnbull on “Change My View”

Julia Galef: Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef and I'm here today with Kal Turnbull.

Kal is the founder of Change My View, a subreddit — that is a forum on reddit.com — created to be a place for reasonable discussion, where ... Basically what it says in the title, people come with a view that they are interested in maybe changing in some way. Or having people show them aspects of their view where they might have been missing something, or where there might be more nuance than they realized. Other people on the subreddit point out those potential flaws and blind spots — and often, people come away with a changed version of their original view. It's a true rarity on the internet.

So, we're gonna talk today about how Change My View works, what makes it work, and how it came to be. Kal, welcome to the show.

Kal Turnbull: Hello, thank you for having me.

Julia Galef: So, how long ago did you start Change My View.

Kal Turnbull: Just over five years now in January 2013.

Julia Galef: What were you doing at the time in your life?

Kal Turnbull: So, I was in my final year of secondary school...

Julia Galef: Which, for us Americans, is?

Kal Turnbull: High school.

Julia Galef: Thank you.

Kal Turnbull: Yep, high school. Basically I was coming towards the end of that year. I had my grades, I had my university offers and so it was a bit of a weird time, bit of a reflective time and so, yeah. It all came about in that moment.

Julia Galef: So it's been around about five years, how big is it? How many people use it?

Kal Turnbull: We have 549,000 subscribers at the moment.

Julia Galef: Wow.

Kal Turnbull: Yeah, it's growing quite fast really. Maybe about 800 subscribers a day on average.

Julia Galef: Tell us how it works. Let's say I have a view that I want to put forth for challenging, what do I do?

Kal Turnbull: Yeah. So you have a Reddit account, and then you submit a post. You check out the submission rules, and then you write up your view. But it's not as simple as that, you do have to go into the background of your view. Which it sounds quite obvious, but I think a lot of people forget to do that in everyday life when they're trying to argue an idea. So that's our rule A, to explain the reasoning.

Julia Galef: When you say background, do you mean like causally how you ended up with that view, or do you mean the reasons why you think the view is correct?

Kal Turnbull: The reasons why, but not just that. We like to get people to explain their background personally, so things they've experienced that might have led to that point. Because we take the opinion that a view on the surface can look the same between two people, but they might have got there for very different reasons, and that can really help when trying to change it.

Julia Galef: Yeah, great point.

Kal Turnbull: So, that's one thing we ask of you. Also just an accurate title that isn't sensationalized — just summarize your view and then submit.

Julia Galef: What's a couple of examples of views that people might submit to the site?

Kal Turnbull: Oh, I mean, literally anything you can possibly think of. We aren't just political, but as you can imagine we've had a lot of posts on politics, particularly American politics recently. But we also take views on culture, your favorite movies or music. It's literally any view you can think of has probably been posted.

Julia Galef: What about views on lifestyle, like “people should stay home with their kids” or “people shouldn't try to earn money, or try to be rich” or things like that?

Kal Turnbull: Yep. Philosophy, the meaning of life, politics, all sorts of things.

Julia Galef: Excellent. So you submit your view, you explain the background behind it.

Kal Turnbull: Yep. And then you press submit and hope for the best. Typically, the people who respond are quite civil, as we have designed the subreddit in that way. The commenters’ rule 1 Requires that your direct responders disagree with you in some wa, so you aren't going to just be agreed with by posting, which is another important thing.

From there, we just ask that you interact and you actually engage with people and talk to them. If you don't, then the post gets removed. Rule E.

Julia Galef: How quickly do you have to reply?

Kal Turnbull: We just want some evidence of conversation within the first three hours, which we feel is quite lenient.

Julia Galef: Yeah.

Kal Turnbull: You don't have to sit there for the whole three hours, you just have to reply to a few comments in that time.

Julia Galef: You have to give a decent signal that you're not just there to troll people or to waste their time.

Kal Turnbull: Yeah, absolutely.

Julia Galef: And then do you respond? So commenters disagree with you in some way, do you argue with them? Are you allowed to argue with them?

Kal Turnbull: Yep. So you can explain why you still disagree with the commenters. We don't have a rule against you disagreeing with people, but what we ask is that you demonstrate willingness to understand, and open-mindedness, which is our rule B. It can get quite complicated, because we accept that not changing your view doesn't necessarily equate to close-mindedness. But we do typically recognize behavior that we deem to be close-minded, and that's just kind of brushing off people without actually responding to what they've said, maybe copy pasting your argument over and over. Just hitting your head against a wall, really.

Julia Galef: And when you say you recognize that behavior, do you openly discourage it? Do you penalize it?

Kal Turnbull: Well, we remove the post if we see that behavior for rule B and that's put on the user's record. So three rule Bs in a row would lead to a ban, a temporary ban.

Julia Galef: What are some of the most important rules that you have for submitters and commenters, or are they the same rules?

Kal Turnbull: So I talked a bit about rule A, explaining the reasoning. That's definitely important, fundamental, I would say. Rule B, behaving in an open-minded manner. They are the two most important. Arguably, rule E comes next with, like I said, having to respond in time.

But probably our most important rule would be rule 2 in the comments, which is "don't be rude or hostile to other users." And we have a very specific meaning here. We aren't referring to what may be deemed as offensive, but just the way you talk to people. So insults and just generally being rude.

It's really important that people don't break this rule because you can see it quite often, if they do it just throws off the conversation and can create a form of backfire effect and make it hard for anyone else to get through to the author of the post after that point. Just changes the tone.

Julia Galef: I also notice a big cultural effect of allowing stuff like that to happen. My sense is that some people are just never rude or hostile by disposition, but a lot more

people are willing to force themselves to be polite as long as other people are forcing themselves to be polite. But as soon as they see people breaking the rules, they're like, "Well I guess it's no holds barred." Or "Well if they're doing it then I might as well too." And then you get this huge — your culture collapses.

Kal Turnbull:

Yeah, that's true.

Julia Galef:

So that's my way of agreeing with you that that seems like a very important rule.

Kal Turnbull:

Yeah, and you'd be surprised the amount of times we get people complaining to us that essentially "they started it." You'd think that argument, maybe people would grow up out of that, but-

Julia Galef:

You'd think they would be embarrassed to give that defense, wouldn't you?

Kal Turnbull:

Yeah, I know, I know. But yeah, that rule is so important. And I think a lot of the time as well people break it without realizing the downsides of breaking it. So we do actually get people apologizing to us sometimes, going, "Oh, it just slipped my mind, i don't normally behave that way, and I can see why it's counterproductive."

If you think about it, these people, they want to change the view of the other person. And so by being rude, it's kind of going against their own interests, because they're less likely to get through to them.

Julia Galef:

Right. You mention their own interests — I noticed on the site you have an interesting incentive system to encourage people to be good at changing other's views and good at changing their own views. Can you talk a little bit about the deltas?

Kal Turnbull:

Yeah, so the delta system is kind of the nice side of the subreddit. I've talked a lot about the restrictive rule side of it, then we have the rewards side, of deltas. Which is basically, if you're familiar with the delta system, you might know that it has a history in mathematics of meaning change.

Julia Galef:

The symbol.

Kal Turnbull:

Yeah, the symbol delta. So delta-T might be change in time, or change in temperature. And so basically quite early on we recognized this is an opportunity to get people to use this delta symbol in the comments in response to people who have changed their view. So just kind of an acknowledgment token.

And so for a while it was just this kind of weird little acknowledgement we did. And then a couple months later I thought it might be a nice idea to introduce a sort of point scoring system where people could collect deltas and they'd have their own delta score, which would then represent their ability to change other people's views.

Julia Galef: Like karma, basically?

Kal Turnbull: Well, I'd say it's very different to karma — karma being the points you earn just by commenting and submitting to Reddit. A delta represents something of a higher quality. It's a genuine human connection. Someone has said you have changed my view, have a delta, as opposed to just these kind of meaningless up votes and down votes.

Julia Galef: Maybe it's a little more analogous to gold on Reddit, where if someone really likes your post they can award you a little gold symbol...

Kal Turnbull: We do often have to tell people that the delta isn't just a super up-vote. Sometimes people will say, "Oh yeah, you've explained this really well, have a delta," and we actually have to step in and go, "Can you tell us why this has changed your view?" And if it hasn't then we remove it. Because we're really keen on keeping that as a pure record of changed views. Which actually turned out to be a really good thing, because it's provided a data source for these researchers to go in and actually study what changes people's views through the delta system.

Julia Galef: Oh, that's so exciting.

Kal Turnbull: Because it provides that token.

Julia Galef: This was a whole area I was going to get to later, but can you tell me a little bit about what research people are using your site for?

Kal Turnbull: Yeah, so the most famous one probably was at the end of 2016. And it was a research paper into just generally the tone of comments, the language used, how does that affect someone's ability to change a view or not? And this got quite a lot of mainstream attention. It was the subject of a You Are Not So Smart podcast episode, and it was reported in the Independent, et cetera.

Really it was a really great paper, and it opened my eyes a little bit to a few things. For instance, it was really interesting to see that if someone puts across their view in a kind of group terminology, so if they say, "We think this, because we've experienced this," they're much less likely to have their view changed. Because they represent a group, and they kind of stick with that group.

Whereas the people who are explaining their views in terms of "I think this" are actually more malleable. Which is really interesting because it kind of conflicts with the idea that you hear about sometimes this idea of the self getting in the way. You know, "I, I, I," when actually in this case, the I represented open mindedness.

Julia Galef: Right. I guess people's selves are often ... They expand beyond the physical boundaries of you as a person. You might be familiar with this phenomemon where sports fans will, they sort of borrow the glory of their team's successes and failures. And they'll say, "We won," or sometimes, "We lost."

Although anecdotally it seems like people are more likely to say, "They lost," when the team loses and more likely to say, "We won," when the team wins. But there is this sense that your glory, your status, your worth as person is determined in large part by the achievements and the status of the group that you're affiliated with.

Kal Turnbull: It's very interesting. We see that.

Julia Galef: Yeah, that's super interesting. So going back to the delta system, is there any way that you reward the submitters for changing their mind, or is it only the commenters who get rewards for changing others' minds?

Kal Turnbull: So this is a really interesting question. It's something we've thought about a lot, because currently the delta system rewards people who are persuasive. And being persuasive typically does have a lot of ... You have to be good in another way. So you can't be rude for example. Sometimes the rudeness might work through pure brute force, but typically being rude means you're not going to earn deltas.

However earning deltas, earning lots of deltas doesn't necessarily mean you are open minded yourself. So it's this interesting thing where we're rewarding the persuasion, we're not necessarily rewarding the open mindedness, because you don't get a delta or a delta score for changing your view.

The position we've kind of come to on this is that surely gaining a new view, losing that bit of ignorance, is kind of a gift in itself.

Julia Galef: I agree in some sense, for sure. But I wonder if the sting of having to be the person up there, exposing the weak underbelly of your view to the internet, and having to concede points... might need a little sugar to help that medicine go down. Even if you know the medicine is good for you.

Kal Turnbull: Yeah, yeah, I know. I guess we didn't want to add any further incentive to giving a delta, because we want to see deltas and know that they were genuinely given out of just pure honesty as opposed to them getting some kind of point as well.

But yet it's an interesting question. I think there is a real braveness to submitting somewhere like Change My View. We get used to the idea that for some reason that having your view changed or being criticized is a weakness, when actually you see these people posting and I think it's a real strength. So we encourage them.

Julia Galef: To me there's almost a kind of counter signaling aspect to acknowledging weak points in your view, or even to just be willing to put it out there for people to criticize.

And by counter signaling I mean: kind of analogous to the way that traditionally people will try to dress in nice, expensive clothing to show that they're high status, but there's another way to use your clothing to appear high status, which

is to wear a kind of scruffy tshirt or hoodie and jeans. And what you're signaling with that is "I'm so high status that I don't need to convince people with my clothing that I'm high status. People know who I am, they know that I'm influential and successful, and so I don't need the clothes."

And so that's kind of a stronger signal, if you can pull it off.

That's often how I think about how status interacts with this willingness to change your mind, or admit fault publicly. If you are willing to do that, and you do it sort of graciously and confidently, that to me is a stronger signal that you know you're competent, you know you're smart, you don't have to defend every single view that you have to prove that to yourself or to other people.

Kal Turnbull: Yeah, there's a real security in being able to put yourself forward like that. Yeah, absolutely.

Julia Galef: Or like, I don't know, men who are willing to wear pink or something without having to conform to every signal of traditional masculinity in order to prove how masculine they are. Something like that.

How often do you see people changing their mind? Like what percentage of the time, roughly?

Kal Turnbull: Well it's hard to give a percentage, but what we do is we add this thing called a flair to a post if it's had a delta given within it. And honestly the front page of our subreddit, I'd say at least half of the posts tend to have the flair. But I can't really give much more detail than that, because it's just such an unpredictable thing.

But it happens a lot more than people would think. I think another reason for that, though, is that we encourage the idea that a change of view isn't just a black and white thing. It's not a 180 reversal.

And people get stuck in this sometimes, they don't award a delta straight away because they think, "Well, I still believe the kind of underlying thing I'm talking about." When actually we see a change of view as a just a better understanding of any kind.

And it helps to literally think in terms of view and sight. So we use the kind of idea that if you've moved even slightly around, then you've got a new perspective, and so you should acknowledge that new perspective.

Julia Galef: I like that. Do you happen to have any examples of what a change of view can look like, that's not 180 degrees?

Kal Turnbull: Yeah, so for example it can literally be, "Oh, I've never heard of that argument before." Maybe you just have a better understanding of the other side's knowledge.

But to get even deeper — although we don't encourage deltas for this, in fact we would reject it — a change of view could technically mean that you see your own side a bit better too. But we don't allow deltas for that because it gets too kind of into rule B territory really.

Julia Galef: Right. Yeah, I mean I find that people with the strongest — in the sense of most actually defensible — views on things, tend to be the ones who have sought out ... They've basically done Change My View informally in real life or in other parts of the internet, they've sought out the best criticisms of their view possible.

And they often still end up holding it, they're just holding a better version of the view than they used to. Their justifications are less flimsy. Maybe they have a better understanding of what the potential exceptions to their view would be when they hadn't thought about that before, et cetera.

So that does seem to me like a really valuable result of trying to have someone change your view is that your view ends up better for it.

Kal Turnbull: Yeah, absolutely. And we acknowledge that, that sometimes by going through this process and coming out with technically the same view, it might seem like a failure of the subreddit when actually it's not. We don't have a reward system for it, but we recognize the value in that. And so yeah, there's that side of it too.

Julia Galef: Have you noticed anything about which commenters are unusually good at changing a submitter's views? What are they doing right?

Kal Turnbull: I think a lot of the time they're just asking questions in a really productive way. So basically the Socratic method. But not in a loaded way, or an obviously leading way. Just genuinely trying to understand first and foremost what's being said.

And I find a really interesting thing here is that in theory, the act of trying to change another person's view and having your own view changed shouldn't begin any differently. You should always start with the questioning, because either you learn enough to change your view or you learn enough to change their view. So it should start the same.

Sometimes, though it might seem to contradict the subreddit, starting with the intention of changing someone else's view can actually ruin that process a little bit. Of course it's hard to get into the minds of our commenters exactly. But I do tend to see a lot of questioning for the successful ones.

Julia Galef: Is it questioning more like... getting a submitter to think about considerations, or situations that they maybe hadn't considered how their view would handle? Or is it more like, "I just want to really precisely understand what your view is, so that I know what is worth arguing with"? Because otherwise you're going down a bunch of blind alleys where you're arguing with a thing that isn't even what they believe.

Kal Turnbull: Yeah, so there's a bit of both I think. I think there's the questioning that is in order to make sure you're not building up a straw man. But then there's the questioning that is the clever kind, which can maybe get out some contradictions through answers. And I have seen cases in the subreddit where just through pure questioning, the contradictory answers created a change of view.

And this is actually quite powerful I think, because the person who's had their change of view feels part of that changing if they are open to it, if they receive it well. Because essentially their answers have opened their own eyes without necessarily being told that they're wrong. So that can be really interesting to watch.

Julia Galef: When you look at the examples where people *don't* change their mind, what do you think is usually happening there? Just in your opinion.

Are they getting good rebuttals that they're not acknowledging, in kind of a motivated way? Or they're ignoring them?

Or do they just submit a really solid view that there aren't good rebuttals to? Like if I submitted, "Gravity is almost certainly real," I'm not going to change my mind about that. So what's happening?

Kal Turnbull: Yeah, I mean examples like that is more on the extreme end. So typically we don't get posts like that, they immediately get pulled down as obviously joke posts.

But yeah, you get the ones in the middle which are really strong views. And sometimes I think people, they might think that posting to Change My View is just something you do to get a view out there, essentially soapboxing, when actually it's not about that. It's for people who generally accept that they might be flawed. They have a feeling that, "Okay, I could be wrong on this."

So the gravity thing, I mean you're not going to have that feeling. So the post is doomed to be a failure from the outset.

Yeah, I mean there's other examples. So sometimes it's because they post a view that they just believe too strongly and they shouldn't have posted. Other times it's this weird kind of mental gymnastics where they've said, "Yeah, I didn't really mean that, I meant this." Or, "Well actually that wasn't what I meant either." And the next thing you know they've actually completely redefined their post without accepting that they've changed their view.

Julia Galef: That's so funny. I wonder to what extent they had just misstated it originally, without realizing that that wasn't what they believed — versus they actually did change their mind, they just don't want to say that they did.

Kal Turnbull: Yeah, it's a tough question, that, because we'd have to be inside their mind to know for sure. But I think a lot of the time it's just a refusal to accept they're

wrong. Because like we talked about, it seems to be a hard thing to do for a lot of people.

Julia Galef:

How would you go about identifying views that would be worth posting on Change My View? So obviously you don't want the ones that are just so clearly correct, like gravity is almost certainly real.

There's some views where I feel genuinely uncertain, like I'm leaning one way, but I can already see that there might be flaws in my view. And those seem like those would be great for Change My View.

But it would be nice if there was a way to identify views where I can't currently see any flaws in them, they seem right to me, but I suspect — just by an outside view, it's unlikely that all the things that seem right to me are in fact right.

Do you have any heuristics, would you say... like, if I know that some smart people disagree with me about this thing, even though it sure seems right to me, is that the kind of thing you would post?

Kal Turnbull:

Yeah, it's tricky. There's lots of different reasons for posting, like we've talked about already.

We typically see — and this is another mod's breakdown and I thought it was really nice — we typically see three types of posters. The first being the open minded, "I could be wrong on this, I have maybe a little bit of doubt."

And then the second is the person that just shouldn't be posting. They 100% believe it, they've essentially written a thesis and it's just they're taking the platform.

But then the third type is tricky, because they're the ones that think they're open minded, but then when confronted shut down.

And so the first two are really obvious, we accept the first and reject the second. But the third, it gets ... Yeah, we don't want to tell them off, because they came here with good intentions and it just didn't work out, but yeah... I don't know if that answers your question, but the kind of ... It's very complex, the reasons why people post, and the varying levels of doubt they have. It's hard to really define it other than those three categories really.

Julia Galef:

I don't know if you have way of observing this, but: Do you think that the people that don't seem to change their mind on the spot, during the conversation about their submission, do you think that they might end up having a more nuanced or moderated view later on because of that conversation? Like, they're just not good at sort of acknowledging they were wrong in front of other people, or things need time to percolate?

I often find this myself. There's an argument that I think is wrong, or I can think of a rebuttal to it so I reach for that rebuttal. And then, I don't know, days,

weeks, later, it comes back to mind and I'm like, "Well, I don't know, I can kind of see that person had a point." Or I notice that I was interpreting their argument uncharitably, I was inadvertently strawmanning what they were saying, and it occurs to me that the more charitable interpretation actually is a good rebuttal.

And that just takes a while. But maybe you have no way to tell if that's happening since these are people online.

Kal Turnbull:

Yeah, we don't know if that's happening. But I would say it's almost definite the idea that we've planted a seed that would later grow into something. I would say that's more than likely.

Another moderator uses a great example. They say that these strongly held views that you've had for a long time are like cargo ships, they're slow to turn around. And I think that's true and so there's a value in the subreddit that exists outside of that immediate conversation.

You also have to remember that there's thousands and thousands of people reading these conversations, so there's undoubtedly been view changes within those people too.

But what we've tried to do, in order to try to get an answer to your question, is we set up our own podcast to try and get some of these people on, maybe a week or so later, to ask them about it and see if they've thought on it some more. That can be really interesting. Once you get them out of the heat of the discussion, because sometimes these discussions are like 300, 400, 500 comments long and it can get a bit full on.

Julia Galef:

I'm curious about how one goes about building a community or a culture really in which these standards of discussion, and the kind of values that you're upholding — like why it's good to change your mind, or the value of asking questions of someone to really understand what they're saying before jumping in to argue with them... How you build a culture where those things are kind of the norms?

So to what extent do you think the fact that Change My View often succeeds is due to just the top down moderation? That you guys are enforcing rules, and you're just really a stickler for deleting things that don't follow the rules?

Versus, the community getting used to "This is how the rules work", and they follow them kind of automatically at this point, and maybe even police each other a little bit by pointing out that someone was being rude, or not interpreting the submitter charitably, et cetera?

So, top down versus bottom up?

Kal Turnbull:

Yeah, there's a lot of community policing, just because of the sheer number of comments and posts we get. There's 25 mods, and it would be really hard for us

to get through everything if it wasn't for the report system. So our users are really good at reporting things.

But the culture we've developed is really interesting because, like you've talked about, we've got that restrictive side of the rules and then the reward side of the deltas and I think somewhere in the middle is this kind of perfect little breeding ground for this kind of culture. And it basically, what it comes down to is: people come in, they see other people earning deltas, and they see that they themselves are not earning deltas, for instance. And they wonder why. And so they'll look into the behavior of those who are, and they'll start to pick up that behavior a little bit.

And this I know because one of the research papers that I mentioned earlier interviewed a few of our users, and one of them talked about this. They came in with kind of combative behavior, and earned no deltas. So they had to adapt and they just followed some of our top delta earners.

So that's the kind of upper side of it. Then there's the lower side of maybe the people who they see that through breaking rules, they're just having no effect.

Julia Galef: Yeah, like wasting their own time.

Kal Turnbull: Yeah, they're wasting their own time — what's the point in coming in and insulting someone? Because you know that can't be healthy for you, and then it's not changing anyone's views. What's the point in that?

And so yeah, I mean we do have to ban people a lot who just will never get it.

But I do believe there's a good strong group of people in there that maybe didn't start out with ideal behavior. We can all fall victim to some of these behaviors sometimes, we're only human. I never claim to be the perfect view changer and persuader myself and I'm sure the other mods agree with that. There's a lot of leniency in the rules as well. I just think it's a really productive environment.

Julia Galef: Can people get a comment reinstated if they fix the problem with it? Or if they get banned can they get reinstated as a participant if they shape up somehow?

Kal Turnbull: Yeah, we've got quite a good system I think. So if someone, they have an insult in their comment, and it gets removed, if they edit it out — then we'll approve it, but we will keep it on their kind of track record. Only because there is a loophole there if we were to remove it. So the damage has already been done. Sometimes they'll edit it out hours later after the damage has been done —

Julia Galef: Does it ever expire though? Like items on your credit report, after seven years or something?

Kal Turnbull: Yeah, we've got a statute of limitations of six months. It sounds a bit silly that we've got that, but actually it's completely necessary because we have long time

users who've been with us for five years, and if we were banning them for behavior they did in year one, then that would just be totally wrong. Because part of the idea is that people learn as they go along. So we want to allow that space to learn. So yeah, we don't punish behavior from over six months ago.

But yeah, so you can edit it out and we can approve it. But we do keep track still.

And so basically we built up this system where a few violations lead to a temporary three day ban. If you come back and do it again then it's 30 days. Come back and do it again and then it's permanent. So really to be permanently banned from Change My View takes a lot of effort. It takes deliberate ignorance of our rules, and just ... Basically the people we've permanently banned just aren't the people we want there, so there's a win win for everyone.

Julia Galef: How have your rules, or your approach to moderation changed since you started?

Kal Turnbull: Well for instance we had to introduce this statute of limitations. We had to introduce the kind of ban procedure of 3, 30 and permanent. So that came about after a long time of kind of seeing how lenient we should be.

The rules themselves, it was an iterative process. We saw behavior we didn't want people to repeat, so we would add it to our rules. For instance rule 2, we're still debating that rule within the mod team-

Julia Galef: What's that again?

Kal Turnbull: And there's that blog post on changemyview.net with a really nice insight into kind of mod discussion on an example of rule 2 behavior.

Julia Galef: Sorry, what is rule 2 again?

Kal Turnbull: So rule 2 is don't be rude or hostile to other users. And basically to just give you a quick rundown, it was about someone who used the words "that's bullshit" to describe someone else's view. And we were wondering if that should be counted as rude or hostile or not.

Long story short, we opted to approve that, so long as it's contained within a more substantial comment, just because there's too much ambiguity there. You can read the post to get a better insight because actually it's a really interesting kind of discussion. We wouldn't encourage people to use those words for someone's view because it's probably not going to change their view anyway, but yeah.

And then there's other rules, like the rule 1, which didn't come straight away — but basically quite early on in the subreddit I realized that people were just agreeing with a post and then that was encouraging people to post their views in order to be agreed with. So rule 1, "Top level comments must disagree in

some way," essentially is what that rule says, and that came about a few months alter.

Julia Galef: Or ask a question, right?

Kal Turnbull: Sorry?

Julia Galef: You must disagree, or you can ask a question, right?

Kal Turnbull: Yep, you can ask a clarifying question. So it doesn't have to be, like I said, it doesn't have to be an immediate disagreement. We encourage the questioning process as well.

Julia Galef: What is a problem that you still feel like you haven't quite solved on Change My View?

Kal Turnbull: Oh, good question.

Julia Galef: Or just a way it could be improved, the way it works.

Kal Turnbull: Yeah, I mean we're always getting better at spotting rule Bs. The fine balance between someone who's just not being persuaded yet and someone who is actually close minded and just there to post their view and to try and spread it. That is something that I think can always be improved. It's really quite complicated and we sometimes have an hour long discussion on whether to pull a post for rule B or not. So there's a lot of learning in there.

I think we're always just updating the rules. Even a couple of weeks ago we added a further clarification.

Julia Galef: What was that?

Kal Turnbull: Rule clarity and encouraging people to understand the delta system.

I think maybe something that could be improved is general understanding of some of the philosophy of the subreddit. It's not just a place to go and argue with other people, we do actually have an underlying philosophy of "it's okay to be wrong, it's not a bad thing."

And you need to understand other people in order to change their view effectively. Shouting and arguing like you might do on Twitter for instance isn't the way to go about it. And so we're always trying to figure out ways of maybe getting that across. But actually just through using the subreddit I think people learn that anyway.

Julia Galef: Yeah, it's often easier to communicate a philosophy or a value just through repeated example, than by trying to define it in words and hoping that people internalize it immediately.

Kal Turnbull: Yeah, absolutely. Yeah, and that's where the delta system is just so great because it's like a reward for that behavior without actually having to tell people to behave specifically in that way.

Julia Galef: Incentives are just so powerful. I don't think I fully appreciated the importance of incentives in people holding on to false views or not acknowledging evidence against their views. I used to just think, "Wow, people hold all these wrong beliefs about politics, et cetera, and they're so bad at recognizing when evidence contradicts their view, et cetera."

And it didn't occur to me until much later: Well, what is their incentive to change their mind about their political view? They don't get any personal benefit from having a better, more accurate view on immigration or tax policy. And on the other hand they get all of these emotional and social benefits from continuing to hold the view that their whole subculture thinks is right and virtuous.

And so the incentives are just kind of stacked against updating, unfortunately, when it comes to especially ideological issues that don't bear on people's personal fortunes.

So it's just really nice and refreshing to find ways to kind of fix those incentives to some extent. To reward people, even just socially reward people, for changing their view.

Kal Turnbull: Yeah, absolutely.

Julia Galef: I'm curious if you've gotten any pushback on your policy of allowing people to post any view, even if it's offensive to people, as long as their tone is not rude or insulting. Has anyone argued that, "Well some views just shouldn't even be allowed to be posted and discussed in reasonable tones, because giving any credence to offensive views is bad and harmful?"

Kal Turnbull: Yeah, I mean we definitely get that. Sometimes people message us saying, why do you even allow this to be discussed?

It's a really interesting thing, because basically we take the stance that we are there to moderate the manner of conversation, not the topic. We feel that if we start moderating the topic, then we kind of become this weird arbiter of truth. We don't want to get into that territory, so we allow our own community to respond to posts they want to respond to and ignore the ones they don't. Because ultimately the best way for a post to die is just everyone ignores it.

And it's also this idea that... okay, say you take out views that someone else deems offensive. Then the ones that are most offensive next, you take them away. And then you take the one underneath that, and next thing you know people are only allowed to post nice, good views. And then the subreddit falls apart because people have to disagree with those nice, good views.

And so our kind of stance is, first of all we don't want to tell people what to talk about. But secondly, nobody wants to change benevolent views.

So it's an interesting problem, one that we definitely get some criticism for. But we're not promoting views by allowing them to be posted. Because even if you took that stance, well, within the post itself top level comments are criticizing the view. So which are we promoting? Are we promoting the post or the comments? Obviously we're promoting neither, we're promoting conversation, open civilized conversation.

Julia Galef: I think also, in my view it also really helps dissolve some of this objection that you guys are ... that the real purpose of the site is to promote this way of dealing with disagreement. And this way of questioning one's own beliefs. And so allowing people to do that with views that are fringe or offensive is ... The reason that you're having those discussions is not to spread [the views] more widely in society, the reason you're having the discussions is to practice these valuable norms. that I think most people agree would make society better off.

So I think having the focus be on the principles underlying the discussions and not the object level topics probably helps.

Kal Turnbull: Yep, absolutely. And as another moderator says, sunlight is the best disinfectant.

Julia Galef: Sounds like your moderators are great. You've quoted several of them, all with insightful points. How do you find your moderators?

Kal Turnbull: Well we tend to add mods who have been solid contributors for months if not years. Most of them have quite high delta scores. They're just very interested, and they love the subreddit, so it makes sense to get them on board.

We recently added five more just because the sheer popularity of the subreddit is becoming ... We need more people to help manage it. But we're on top of things for the time being.

Julia Galef: Two questions: one, what kind of people are the participants in general but especially the moderators? I'm curious, are they young, old, are they in quantitative STEM fields or humanities or neither? And then also, is there a lot of overlap in your object level views, you and the moderators? Like do you tend to share some ... You're all atheist and you're all progressive et cetera, et cetera, or do you have a wide spread?

Kal Turnbull: Well the really interesting thing about this is we kind of have the purest of ways of adding people to our team. Because we put out a post asking for applications, and basically all you have to do is, first of all, be a positive contributor. But also answer five questions. And none of them are personal questions.

So actually we don't know anything about the person other than maybe some of the topics they've talked about, until they become comfortable on the team.

And even then, out of the 25, not all mods have told everything about themselves to the other mods. So there's only a handful I could say I know and understand. But there's definitely diversity of thought in the team for sure. Absolutely.

Julia Galef: So Kal, I'm going to let you go in just a minute, but before we wrap up today, I wanted to ask for your Rationally Speaking pick. So this is a blog or article or book or could be even a movie or a play. Something that has influenced your thinking in some way — changed your mind, say, or gotten you interested in a topic that you've been studying since. What would your pick be for this episode?

Kal Turnbull: So I would have to say a book that I'm just about to finish, I've not quite finished it, but "The Righteous Mind" by Jonathan Haidt. It might be an obvious one to some people that I've chosen this, but it really opened my eyes because it talks a lot about these things that I've been talking about without me having fully understood them. Because I come from a background that I haven't studied any of this stuff in an academic way, I've just got this pure experience of the subreddit and what it means to change views in that regard. And "The Righteous Mind" has just really kind of nicely opened my eyes to some of these behaviors and why people behave the way they do.

Julia Galef: Like why they hold the views they do, why they do or don't change them?

Kal Turnbull: Yeah. I mean particularly the idea of the elephant and the rider. This idea that our intuition leans and our rider can often, all they can do is try to guide it. What a great way of expressing something that I kind of just have this idea of but didn't fully understand. It's been a really interesting book to read, so I would definitely recommend it.

Julia Galef: Excellent. Well we'll link to "The Righteous Mind," we'll certainly link to the Change My View subreddit. And the podcast that you mentioned, how do our listeners go find your podcast on Change My View?

Kal Turnbull: Yeah, so we've got a separate website called changemyview.net, which is part blog, part podcast, and you can listen to the podcast there. So yeah, that's where you find that. The URL to the subreddit is [reddit.com/r/changemyview](https://www.reddit.com/r/changemyview).

Julia Galef: Excellent. Kal, it's been such a pleasure having you on the show. Thanks so much for joining us.

Kal Turnbull: Thank you, Julia. It's been really great.

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