Participant #7: So, why did I choose this prior? My memory of this was that the maximum number of pumps was 120. I put a prior that didn't have too much weight over the maximum. Looking at this now, I probably could have built it a bit to the right, but I remember the heavy tail being a bit sort of different with ... I went with the idea of that that tail's still quite heavy, so there's still a lot of probability mass to the right of 100. So it shouldn't get in the way too much. The other thing that was in my mind was that this was one of two parameters, so this was for Setting 0.

Participant #7: But for Setting 1, there was another ... So you multiply this by another number. I was a bit nervous about that right-hand tail, knowing that there was a high maximum.

Interviewer #1: If you stay on that, so the prior that has been visualized is the default setting, that's not the one that you selected? You chose a normal-

Participant #7: Oh. Did I?

Interviewer #1: Yeah.

Participant #7: [crosstalk 00:01:15].

Interviewer #1: Yeah. You chose the normal 3.6.

Participant #7: ... Find it. Oh, yeah. Okay, that makes more sense. That's got a lot more mass around the 120. I mean, I'm not the world's biggest fan of this big hump down the zero. It's an odd piece of prior information that I think that there's a median sitting about here in the distribution. That's sort of the wrong place. I would want the median to be more up here, but the shape of the log-normal prior just kind of gives you what it gives you. That was essentially ... My thought process was basically I needed to not go through prior above 120, but I'm also going to have to multiply this by something.

Participant #7: So this was the closest I could get. That make sense?

Interviewer #1: Mm-hmm (affirmative). Yeah.

Participant #7: And the parameter, or should I move on to the next one?

Interviewer #1: I guess if you scroll up, I think you were talking about the description that we provided. See, if you scroll up, there is that.

Participant #7: Yeah.

Interviewer #1: So the description button. I was wondering, did you consider any other information that was over there while choosing this intercept parameter?

Participant #7: It was pressed on, so that we were modeling the log mean. So those were two pieces of information. I actually don't remember if I knew the fact that the average is between 20 and 40, although looking at this prior, that's where I put it. That's ... I definitely didn't use the wide standard deviation, because I've got no real way to put this in. Also, I always worry about cleaving too closely to meta-analyses of past studies, because as we all know, it's better to be wrong ... It's better to be slightly wider with the prior than slightly too narrow.

Participant #7: So, I used some of this information, but not all of it. I was sort of much more coming from a weakly informative point of view where you work on sort of bounding...containing the sort of the probability density in a sensible part of the space. [inaudible] to know what the actual average in these two conditions in this prior ... Well, these combinations of priors are. I suspect they're way bigger than these numbers.

Interviewer #1: Well, why do you suspect that?

Participant #7: If you take that parameter and multiply it by that parameter, I think you're going to get a number that is bigger than 40.

Interviewer #1: I see. Okay. I don't have any questions regarding intercept. I2, do you have any?

Interviewer #2: Just if there is any other information that might have been helpful here or any aspects of the interface that were not useful.

Participant #7: The information that I didn't actually know what it meant was, what is a weighted standard deviations in this context? In terms of this was pretty good, although the piece of information that is missing is the sort of the marginal distributions under the two conditions. I mean, it's nice when you've got a binary variable. You can sort of put it in with both. It's just a really nice setting to do that in, because that would be an easier way to see ... I mean that's less important for this one, because this is a directly interpretable ... The intercept is a directly interpretable parameter.

Participant #7: The difference parameter is a multiplicative difference, and that's the log difference, and that is harder for me to get a feeling of. So, it would be kind of cool to see what this ... I shouldn't point with my fingers, what this choice of prior and that choice of prior did.

Interviewer #2: What they implied as the marginal distribution of those condition.

Participant #7: Yeah. In a special case where there is ... Like it's easy to do that. So, if you had a continuous covariate here, then you wouldn't be able to do that, obviously. But for a sort of a two-level factor, it's quite straightforward to do.

Interviewer #2: Right.

Interviewer #1: I see. Cool.

Interviewer #2: I think that's it for that section, for me.

Interviewer #1: Okay. Let's move on to a different one. Could you talk a little bit about how you built this prior?

Participant #7: Forgot what it was. N(-0.56, 0.27). Yeah, it got really peaky. I remember that. I mean, essentially I wanted this, one of those situations where you want some amount of expert knowledge, like of what an effect size realistically should be. I mean, how different should these conditions be? I thought something like a one and a half times was going to be about as big as it could get. [inaudible] was that obviously I am not particularly expert on this information. I wanted that quite heavy tile, so it has quite a heavy tail. I mean, I don't like that it sort of is asymmetrical around one.

Participant #7: I would really like this to be centered at one, and between about, with a scale parameter, a sort of .5 would be sort of my ideal version of this prior. But for some reason ... Oh, look. I could do it by doing that. What the hell is wrong with me? This is the metric. Got you. Sorry, I'm working my way through some feelings. Yeah. The asymmetry here is a bit sort of weird to deal with. This prior is saying that I think the effet is much more likely to be smaller when I move it over here. I'm not completely sure that's true. It just happens to be like a weird consequence of the log transform.

Participant #7: Yeah. I would probably, looking at it now, have a prior that looked more like this. Sorry about that, but even then, the asymmetry concerns me and really even this two times here. I mean, this is going to give you for the most part, like 50% of the mass is definitely above 40, which is the upper bound on the knowledge you've got at the top. There's a big chunk of probability, like 5, 10% of the probability that's above 80, which is the value of that parameter being at its mean and then ... Anyway, it's really hard to think this through. I'm very bad at multiplying densities together in my head. Sorry.

Interviewer #1: No worries.

Interviewer #2: That's exactly the kind of thing that we want to know about.

Interviewer #1: I guess my followup question would be, if you are unconstrained by the limits by this box, the grade that we provided you, what kind of a prior would you have chosen?

Participant #7: With the information that I've got and the fact that it has to be sort of a log-normal or a log-T, probably one that looks like this, to be honest. Maybe I would make it look tighter and just sort of trust those heavy tails to do some of the work for me. That really does depend on ... I would want to talk to somebody who does these sort of risk aversion studies about how much more risky they expect people to get between the two conditions. Because there really is an effect size question here. As a basic guess-timate, I do not think that these effect sizes would be huge.

Interviewer #2: If you asked someone who was a domain expert here, what would you ask them?

Participant #7: Questions that you'd ask for a prior analysis. Things about, what do you think is the smallest effect size that you think is meaningful? And what is an effect size that you think would be fantasy? Sort of try and interpolate between those two lines. Again, the heavy tail on the Student t lets them be a bit wrong on the fantasy side of things without hurting us too much, which is sort of nice. Heavy tail in these situations.

Interviewer #2: Anything else in this context that, similar question to before, that is missing that would be helpful?

Participant #7: As before, it's just like this parameter is only interpretable relatively. So, that's just challenging. It's a classic problem with these sort of log linear models.

Interviewer #1: While interacting with this, was there anything else that you found difficult or challenging?

Participant #7: This was a good way to do it. I really liked that it was on the effect scale. [inaudible 00:12:52].

Interviewer #1: I think we lost your audio.

Participant #7: Sorry. Can you hear me now?

Interviewer #1: Yeah.

Participant #7: In this particular content, I probably wouldn't call this intercept parameter. I would call this effect under condition zero or something like that, mean under condition zero. Similarly, I would make this mean under condition 1. Have a prior with mean, under condition 1, just because [crosstalk 00:13:30].

Interviewer #1: I'm sorry, go ahead.

Participant #7: Just because it's more informative labeling for people who are trying to put in expert information, because intercept is one of those words that we all like, because we all know what it means. I quite like the box. I think there was ... Maybe in this second, this difference parameter, maybe there wasn't enough room to go and make a really peaky prior, but that's a weird function of this context where you might want a really peaky prior. But its hard to tell with the information, except [inaudible] between the 20 and 40 is about twice.

Participant #7: So, maybe that is peaky enough. Kind of hard to tell, to be honest.

Interviewer #1: Was there any other information that you wanted, besides the things that you already mentioned, like effect sizes from experts and stuff?

Participant #7: Honestly, thinking about it now, probably this information here about the average number of pumps in the prior studies. It's not phrased like that.

Interviewer #1: I see.

Participant #7: The not average across conditions version of this. For condition 1, it's between 20 and 40, and for condition 2, it's between 30 and 40 would be the information that would be needed to work out how big the mean difference parameter should be.

Interviewer #1: I see. So, I guess we're going to move on to the next page.

Participant #7: [inaudible 00:15:30].

Interviewer #1: Cool. This one takes a while to load. Okay, it's loaded now. Again, like this page consists of three visualizations. The first one is the density on the parameter scale. This is basically the density of the parameter. The second one is on the response scale, which you've already seen, and the third one is the prior predictive probability density. First, we'll look at the first one, which is the density on the parameter scale itself. Looking at this visualization, do you think this information would affect your choice of priors in any way or how would you use this information in any way? Could you contrast this visualization to one of your response scale?

Participant #7: [inaudible] basically, because I'm pretty good at my logarithms at this point, but I don't think it's a particularly good [inaudible 00:16:46]. Its just sort of ... Even doing here, I'm sitting here thinking what is e^5, what is e^6? I have that information somewhere and if not, I would just look it up, but I would essentially be converting it to the proper scale, because otherwise it's quite different. I think you could do, if you provide the logarithms of these numbers, it may be easier to do, but really it's just incorrectly getting through that.

Participant #7: I don't love this. I mean, after like years and years and years and years and years and years and years of setting priors like this, I have realized that was stupid. I could've saved myself a lot of heartache. It's just exponentiating.

Interviewer #1: So ... Go ahead, I2.

Interviewer #2: So, that's actually a really interesting question. How have you set priors like this in the past?

Participant #7: So, in all kinds of ways. I mean, I've actually done a lot of research on prior distributions, because I've done a lot of software development. I've been involved in a lot of that and we've had to come up with some sensible default priors and all of that sort of stuff. I went from conjugate gammas, all the way through to things that are like large Gaussians with massive variants and all of those things and getting tighter and tighter and tighter and tighter. Eventually, I realized that almost all of my really annoying problems were both my priors, because they're one of those weird things.

Participant #7: They don't matter at all, unless they do, and when they do, they really, really, really ruin your day. Yeah, I just sort of ... I think when I was in England, I went to a subjective Bayes conference. It's not a thing I necessarily recommend, but they set me right about this. They said, "You can't set a prior on something that is not observable, or at least on an observable scale." I'm like, "That seems sensible," but I felt quite dumb, because I'd done a lot of stuff that was not that. Yeah, it basically came around that time or around when I was working on this paper on setting sort of weekly informative priors where we realized that we had to set some sort of scale and then, what can you ask people for?

Participant #7: Well, you can't ask people what the log mean is. You can ask people what the mean is. I mean, the log mean is, but it's sort of a weirdly indirect question. Then, we came out of having to deal with people who weren't sort of statisticians ... Mistakes.

Interviewer #1: Cool. I was just curious almost, if you are not limited by the choice of the distributions that we had given you, like Student's t versus normal, what kind of distributions would you use?

Participant #7: My priors on my intercepts are always normal, mostly because I'm lazy, but also because in some sense, that's not as silly prior. Like for positive data, sort of a natural way to average things, the geometric mean, which means that you would want it to be sort of flat and symmetric on a log scale. It's not a silly thing. When you exponentiate it, you suddenly realize that you're making some assumptions about skewness, which you may not love. [inaudible] is a normal. If I was putting my parameter on a response scale, I may be tempted to use a truncated normal on the response scale or something like that, but [inaudible] for convenience reasons.

Participant #7: It is deeply inconvenient to work with non-standard distribution sometimes.

Interviewer #1: Okay. I guess we can move on to the prior predictive probability density.

Participant #7: [inaudible 00:21:40]. Yeah. Okay. The prior predictive distribution or change using the priors you have, which generate that and incorporates information with all the parameters, not just the [inaudible 00:21:56].

Interviewer #2: I think we lost your audio for a second there, [inaudible 00:22:10].

Interviewer #1: Yeah. I think we still ... We can't hear you.

Participant #7: [inaudible 00:22:20].

Interviewer #1: Yeah, somewhat.

Participant #7: Sorry. I don't know what happened. My phone died. Well, we're back again.

Interviewer #1: Okay.

Participant #7: Sorry. I was just reading this text. Okay. Prior predictive distribution tells you what the data number of pumps might look like based on your assumptions there before looking at the data. Visualization shows the prior predictive distribution, the density of the number of pumps in each conditions, the 20 hypothetical experiments seem the sensible priors we chose for the mean difference parameter beta.

Interviewer #1: Yeah. So we fixed the mean difference parameter and we just ... At zero, and then we generated these prior predictive distributions.

Participant #7: Okay.

Participant #7: Question when I look at this graph is, why is this multimodal?

Interviewer #1: I think that we used kernel density estimation to generate the density. So, if there's like even a slight probability over there at the tails, it's giving those [inaudible 00:23:34].

Participant #7: [inaudible] always happens. Right. [inaudible] on this, I am very struggling tied to the fact that this axis keeps moving on me, which is driving me lightly mad. That is the thing I am not completely in love with. Secondly, get that bulk to the right of ... Yeah, I think to the right of a hundred. [inaudible] that's to the right of a hundred. It's really hard to tell. KDE's are awful for densities from low sample sizes, because this was 30 pumps per experiment, 20 experiments?

Interviewer #1: Yes.

Interviewer #2: Mm-hmm (affirmative).

Participant #7: Size is too low for a KDE to be like a reasonable way of estimating a density and you're just seeing artifacts on top of artifacts, which makes it a bit difficult. I kind of like that what's coming out of here is maybe that the constrictive case is a bit further to the right. That's going to depend on beta, but that's like ... In theory, I like this plot more than I do in practice.

Interviewer #2: What do you mean by that?

Participant #7: A big part of that is this X axis that is just, I mean, ludicrous. This goes up to 2,200 is when the bit of the distribution I care about is all being squeezed into here. That is a problem. The artificial wobbles are a bit annoying as well. I mean, it's worth saying that this tail goes further to be sort of having like a ... truncate just get rid of this, but like a truncated value --- say the biggest values are sort of 2,200 and whatever, but I mean, tails go out forever. That's kind of the point of tails. I don't know what [inaudible] shit. Sorry. It's just a bit awkward that I can't choose this X axis and just say, "Actually, I want this to be fixed."

Participant #7: If I make this normal, a lot of these problems go away. You have far fewer outliers. A prior like this is well centered sort of in the area where you think the probability mass ... Well, where the expert information says the mass is, but there's enough room for every feasible value to get through that. I'd pick a prior, something like this. Which is different to the prior that I picked just based on a single parameter. I would probably do this as histograms rather than kernel density plots, just because ... Like overlay histograms, just because they're much stabler and you don't get this wonky nonsense.

Participant #7: Also, because the structure of the experiment is that these are sort of exchangeable. There's no drift term or anything like that to take into account that this is the 30th experiment and I'm bored. Then you don't need to plot 20 of these histograms, actually giving ... I mean, I would honestly just plot the distribution of this data. So like histogram of a couple of thousand draws from the constrictive and a couple thousand draws from the expansive and put them on top of each other and say, "Hey, does that look sensible?" Would be the way I think I would visualize this, because of that exchange-ability, which just makes life so much easier.

Interviewer #2: Makes sense.

Participant #7: The wonky stuff that's happening over here is a bit ... I feel like if I was not someone who knows how kernel density estimates work, I would be very, very concerned about these. Whereas now I'm not concerned about that, because you said it's a kernel density estimate.

Interviewer #1: Right.

Participant #7: That's sort of my only big problem with that, but yeah, it's interesting to see how different the prior I set was. Yeah. It was actually a pretty ... Like the right wall. [inaudible 00:29:22].

Interviewer #1: Okay. I guess, is there anything else that you'd like to contrast with the visualization and the other two visualizations?

Participant #7: [inaudible] the things I don't like about the actual sort of implementation of the graph. This is like much easier to work with. It is much easier to translate sort of the description information into like a graph like this. It wasn't too bad to do it here, but Chrome has a very bizarre shortcut apparently that really likes going back a page.

Interviewer #2: No worries.

Participant #7: It would be interesting, just for my own personal curiosity to see what happens if I do this. Actually, that's a prior I don't mind. I don't mind. Not the one I went with originally. The one I went with originally was like here or somewhere, wasn't it?

Interviewer #1: Slightly lower. Yeah.

Participant #7: Yeah, whereas this is actually [inaudible] here, whereas when I was like squiggling around the other day, I went through that part of the space and just sort of kept on going. Sort of interesting that this shows much more clearer that that would've been a good prior, because when you put the prior down here ... Yeah, it's all a bit normal, [inaudible 00:31:23]. Yeah. I mean, there's some overlap between these two things, but yeah, I have a strong preference for this one, just because it makes me feel less awkward about setting the beta.

Interviewer #1: Cool. I don't have any more questions on this page.

Interviewer #2: This is great.

Interviewer #1: Cool. Let's move on to the next page here. So yeah, this was the last page, is the same as the previous page, except that we're looking at the iterations for the mean different parameter or the beta parameter, I guess.

Participant #7: Mm-hmm (affirmative).

Interviewer #1: Again, we look at the first one, which is the density of the prior on the parameter scale. How do you use this information if at all, and like how would you contrast it to the other?

Participant #7: Sort of like this is slightly easier for me, because I know that E to the zero is 1, so I know that this one needs to be centered to the zero and then I just transform the quantiles. I know that two is out about here, so plus or minus 0.7 or something. It's actually this happens to be a logarithmic scale that I'm pretty good at visualizing. That's probably have to do with the fact that I've done a lot of this. Anything else? Actually, that's like a relatively easy prior for me to set in my head. [inaudible] like a normal ... Well, that wasn't the normal zero one to be honest.

Participant #7: The default prior to something like ... I think probably something like this [Normal(0, 0.4)] Is more reasonable, but if I was just running through an ordinary analysis, it would be a normal zero one. Well, until I looked at the sort of the data generating distribution and saw that that was a bit weird.

Interviewer #1: Okay. So again, just to confirm, you would choose a prior ... In this situation, you'd choose something tighter than the upper bounds of this one?

Participant #7: Yeah. I mean, I'd probably choose something that was about ... Let's actually do it. A lot of ... There's still a bit of mass plus or minus 1, but plus or minus 1 is a big ... Plus or minus 2.7 is a big times ... Is a big change, so I would be surprised to see a change that big and [inaudible] that was that big, that had that much of an effect. I just don't think they exist very often.

Interviewer #1: Yeah. Okay. Cool. Let's move on to the prior predictive one.

Participant #7: [inaudible] and sensible, but-

Interviewer #1: It's the same thing, except that this time we fixed the alpha at 3.5.

Participant #7: [inaudible] nothing.

Interviewer #1: Sorry?

Participant #7: Knowing, as I do, absolutely nothing about people and risk aversion, I would not put a prior that put these two ... I would make constrictive and expansive approximately exchangeable. I would center this prior at zero. So, center the multiplication at one. If I knew more about the problem, I might be more willing to do something like this or something like that. That really would come from sort of if that information, the description had some more information in it about splitting the two conditions up. But if I was going to do something as bold as say this, where I'm saying the expansive condition leads to much less clicks, I would probably shove a Student t in there and sort of hide some of my ...

Participant #7: Like give myself a little bit more room to be wrong. Heavy tails save you a lot in these situations. For a model like this, the cost of having a quite heavy tailed prior isn't that big, because the model is relatively simple. The model gets more complex, you don't want to have these hugely heavy-tailed things lying around to just land the sort of tall grass and ruin your day, but yeah. I mean, if I happened to think that this was a possibility from some sort of expert information, I'd probably go with something like this with the the. Otherwise, it would be normal and it would be scaled appropriately to try and like ... Tail that didn't completely preclude an experiment where somebody pumped the maximum number of times.

Participant #7: Because there are always the annoying people who just want to press that button. So probably go further down than in this particular box lets me just to pull that right tail a bit further to the right. That's probably a function of how you set the alpha parameter.

Interviewer #1: Right. I think ... I mean, in this case, I think one issue is that we've fixed ... While generating these samples, we just fixed the alpha parameter, so there's no variation in it. I think I've been noticing here you set the scale at one, whereas previously you set the scale at ... Like at the complete of ... Like 0.2. could you talk a little bit about that?

Participant #7: [inaudible] zero for that one. No, that one.

Interviewer #1: No, the-

Participant #7: [inaudible 00:38:36].

Interviewer #1: No, the [inaudible] the scale of the prior.

Participant #7: The previous settings here for this effective prior on the beta.

Interviewer #1: It was in 0, 0.2. I think ... Not in the survey, but in the first stage of this.

Participant #7: Sorry, up here. Cool. Yeah. Yeah I did, didn't I? If you look at that here, that's a very tight prior. I mean, that's really, really saying that I'm not seeing more than 60 or so pumps, and while I might believe that that's likely, I can't preclude somebody pumping more than that. I try really hard to make ... Like if I've got an upper bound like, this to make it ... I would be less worried if it was a t-distribution cause then there's like a longer tail out here. Also put that to 0.4. [inaudible 00:39:51]. I mean, that's still like ... It's really tight when combined with this alpha, and yeah, I mean, I would ...

Participant #7: It could come out when you do this and then you do your prior sensitivity checks and whatever, that your posterior shrinks strongly sort of somewhere like here, somewhere around 35 or 40, in which case, whatever. The fact that the prior isn't giving you much mass to the right of 60 isn't really a big deal, but I would worry about my prior concentrating sort of around 50. If that happened, then I wouldn't trust the right tail of my estimate. Just from sort of a weekly informative framework, this is quite ... Put it back to normal, because it's not a certain normal. This is quite a strongly informative prior. Doing these dynamic axes, you're never getting anything above 70 in 20 experiments of 30 pumps.

Participant #7: You're never getting more, in 1,200 draws, you're never getting above 70 something, which is a concerning number.

Interviewer #1: Interesting.

Participant #7: Yeah. The assumptions that I made when I sort of set this prior like this were much more ... Implicitly had some assumptions about how the prior alpha was set, that probably aren't followed here.

Interviewer #2: Right.

Interviewer #1: Right. I guess like in this realization, are you even considering what the prior on alpha is?

Participant #7: Well, the prior on alpha is whatever the green lines are.

Interviewer #1: Okay. I see.

Participant #7: Just [inaudible] parameterization. We wouldn't have set that prior on alpha, so then I have to set a ... Sort of adapt to that when I'm saying the prior and beta.

Interviewer #1: Great. Okay.

Speaker 4: [inaudible 00:42:31].

Interviewer #1: Cool. I don't think I have any other questions then.

Interviewer #2: Yeah. I think that is ... I don't have any other questions here. Do we want to just move on to the last couple of questions?

Interviewer #1: Sorry. I just had people drop in. Yeah. I guess ... Yeah, I don't have any other questions regarding this. This is very interesting, then what you've talked about is very insightful. I guess the other questions we had was sort of, what other information do you consider when choosing priors for models like this or any other models?

Participant #7: The other piece of information that isn't particularly relevant here, because the model's pretty simple, has to do with complexity. Basically, it's really when I choose a normal or a the here, I mean for this particular model, it doesn't really make a difference. If you sort of had an extra term with a Gaussian process or a spline or a mixture component or something else, then suddenly is the complexity of your model for this sort of mean grows, the more dangerous if it comes to have quite heavy tails. So knowing what's in the rest of the model would sort of inform what sort of prior I was willing to put on... Like what sort of tail I was willing to have on these sorts of parameters.

Participant #7: That's maybe the big thing. I mean, not really. It's really about what the realizations look like, what a sensible mean is going to be. I'd spend less time thinking about the fact that the mean of a Poisson this, the variance of a plus on that, I probably should, but there we are. Awesome. Good. The really strong mean, variants relationship is a bit awkward here, but that's life.

Interviewer #1: Okay. I guess [inaudible] to my final question. Is there any other ... Can you think of any instances in your past when choosing priors that like ... What kinds of information did you consider and did you have any challenges in determining those priors?

Participant #7: So, one time that we got it extremely wrong and it was with exact same model, but with a negative binomial. So there was overdispersion parameter. Basically when we set the original default prior INLA, we accidentally forgot which direction we'd parameterized this in. So where the big was more over dispersion or less over dispersion. I don't remember which one we did, but we got it backwards. I mean, this happens a lot, people putting priors on variances instead of standard deviations or priors on precision, which are extremely difficult to interpret. So, parameterization really matters. Yeah, there are no traps you can really fall into, except possibly this coding of condition as zero one makes alpha really interpretable that can lead to some problems if ...

Participant #7: If the two conditions are massively different, then beta might have to be a very big number. You can't really keep ... Alpha and beta aren't necessarily on the same scale for this type of setup. If you code the condition as plus or minus 1 or plus or minus 1/2, then alpha and beta are on approximately the same scale or like a little bit more interpretable. It's sort of ... That's a difficult problem. It gets worse when you've got more stuff going on in this mean. So once you start getting all random effects and mix effect models and multilevel models, all those sorts of things, it can all get a little bit much.

Participant #7: Then thinking about the parameterization carefully is kind of important. But for this particular type of example, the only massive challenge ... I mean, is the logarithm always, but also that beta is a multiplication parameter or even beta is a multiplication. Those are the two challenges. It's inherent in this type of model. You can't really fix that, but they get worse as the model gets more complex.

Interviewer #2: With respect to that one-

Participant #7: Sorry.

Interviewer #2: ... The negative binomial example, how did you actually realize that the prior had been set incorrectly?

Participant #7: Really didn't fit, and it took ages to work out, because it was like a piece of software after doing sort of approximately inference. It was like it wasn't clear whether that algorithm didn't work for a negative binomial or whatever it was. But then you do sort of standard prior posterior like checks. You take a look at sort of prior variants, there's posterior variants. You look at where the densities are relative to each other and you sort of ... It was obvious that we were like right up against the boundary. That it was trying really hard to have a very different value of that overdispersion parameter than we were letting it have.

Participant #7: Yeah, we'd sent it around some quite large overdispersion. I think ... I remember it was like, we were doing some fishery stuff with the fisheries people and there was some ... Literally, they had to decide whether that fish stock was going to be overdispersed or not. It was going to be overdispersed, then the negative binomial would fit. If it was going to be Poisson ... If it was not going to be an overdispersed or not very much overdispersed, then putting a plus on would just do much better. From that, we eventually realized what had happened. But yeah, that was an embarrassingly long cycle, because you never think it's the priors [inaudible] everywhere else.

Interviewer #1: Right.

Participant #7: Out of curiosity, is this scale parameter, is that the standard deviation or the variance?

Interviewer #2: It's the standard deviation, isn't it?

Interviewer #1: Yeah, yeah. It's the standard deviation.

Participant #7: [inaudible] information that should be on here somewhere. It might be, I might have just missed it. That's what happened with that. I mean, it's happened multiple times. I mean, any like marginally complex model, we've at some point mess up the priors. Gaussian process is ... Got a lot of time doing Gaussian processes wrong. I spent a lot of time doing like spatial random effects type models very wrong with ... We've inched towards this. It's quite weird how ... Like the ways in which getting the prior wrong shows can be quite subtle, although it gets less subtle when you start looking at predictive distribution like prior and posterior predictive distributions and stuff like that.

Participant #7: It shows up pretty fast that like that's just weird. You're like, "Of course, yeah." [inaudible] sorts of experiences I got varied sort of evangelical about [inaudible] these things. I will say, "Have you tried the prior predictive density?" [inaudible 00:51:13].

Interviewer #2: Okay. I think that's it, unless you have other questions.

Interviewer #1: No, I'm good.

Interviewer #2: No? Yeah. This was very, very helpful. Thanks for spending the time doing the interview and the survey and also for encouraging other people to do the same. Yeah, this was very helpful. [crosstalk 00:51:42].

Participant #7: Really excited about this. I'm really looking forward to finding out what you get out of this, because it's really, really, really interesting.

Interviewer #2: Yeah. Well, I mean, we will ... When we write it up, we'll obviously tweet out the paper and release whatever we get from it. I mean, we're also ... These interactive visualizations and whatever further iterations that we do have them, we're trying to build into a toolkit R, so you can do this stuff with like brms and rstanarm.

Participant #7: Yeah, it would help a lot.

Interviewer #1: I think the [inaudible] like for simple models like this, it still works. Like you said, multilevel models, when there are multiple variables interacting with each other, it becomes a mess. These visualizations are harder to interpret.

Participant #7: It really does, but I mean some of that is like a function of the parameterization being like a little bit wrong. For a multilevel model, it's sort of wrong to put in different variants parameter in ... Like an independent variance parameter and each thing, because like what you actually want to control is the variance of like the linear predictor. Like the variance of the sum. You don't really have any feelings about the individual right down in those variances, except sort of relatively. You can make some of this easier by doing stuff like that, sort of nonstandard parameterizations of the models. Actually, rstanarm does that. rstanarm is a mysterious black box and no one quite knows what it does, but it does that.

Interviewer #1: Yeah. I think I was trying to ... This works better with brms. I think I was strange when I was trying to build the version with rstanarm, it standardizes ... I think it standardizes the intercept and it does some other stuff, which made it ... I mean, complicated to work within this paradigm, but I think it helps in other ways.

Participant #7: [inaudible] to things that are sensible, but would be more sensible if they were documented better. With open source software, you fight sometimes losing battles against developers, Ben is not the biggest fan of documentation.

Interviewer #1: Okay.

Interviewer #2: There you go.

Interviewer #1: Thank you so much again for doing this, taking the survey and doing the interview. This has been really helpful.

Interviewer #2: Thanks.

Interviewer #1: Bye. [inaudible 00:54:41].