Interviewer #1: What was the information that you considered and how you went about choosing this prior?

Participant #9: Got you. Can I look at the previous section to remind myself?

Interviewer #1: Yeah. I think you need to clear. Okay.

Participant #9: There we go. So I just read through the terms of the experiment. This is to try and glean as much domain expertise since this is certainly not something I know about. It talks about how each pump... So there's a maximum size, so 128 pumps. And the explosion point. What the optimal strategy is.

Participant #9: Then I looked at the prior study, so understanding that this is kind of the expected range. I look at the prior studies and the information from the meta-analysis was helpful to... So essentially, previous posterior has become my new prior. So the average number of pumps should be somewhere between 24 and 44.

Participant #9: I think when I select the prior, and the prior is on the number of pumps, right?

Interviewer #1: Right. For the intercept.

Participant #9: Right.

Interviewer #1: And that is on the log scale.

Participant #9: Got you.

Participant #9: (silence)

Participant #9: Let me see what I did. That was normal. I guess there's no way to type that in here to see exactly what I did.

Interviewer #1: Yeah. I mean, it's approximate. So if you go roughly to the area, more to the upper left, then you'll see.

Participant #9: Yeah.

Interviewer #1: Yeah, something like this.

Participant #9: I'm not sure if I followed the log scale. That could have been something I didn't see. But in general, if I have information about what we should expect in terms of previous results as well as what the range, the max sort of is, using the visualization, I tried to make sure that we were... If you can't go over 128 pumps, then you shouldn't be able to have a prior that that's possible. So that was one piece.

Participant #9: Making sure that the range, the average from the meta-analysis was within the middle of the densities for both conditions. I'm not trying to impose any prior information that these should be different, or that should be come from the data.

Participant #9: So my prior belief should be these are essentially the same. Then the experiment should allow that to tease out. I think those were the three sort of thoughts in my mind as I tried to set the prior. Maybe incorrectly, if I didn't follow the log scale.

Interviewer #1: I mean, that's fine. Just to make sure I'm understanding, so you used the widget and you moved it around until you saw a distribution that looked right to you?

Participant #9: Yeah, that looked right in terms of those three conditions. Where the average was within that range, where there wasn't a probability or high probability of it being something that was beyond the number of pumps that there could be, and that the two conditions had essentially the same density.

Interviewer #1: Mm-hmm (affirmative). I guess if at any other point on the interface, if you saw a similar visualization, would you have gone with that? Or was there any specific reason for choosing the mean or the scale parameter besides what you've described already?

Participant #9: Yeah, I think it's those three things, because that was the information that was given, right? So-

Interviewer #1: Yeah.

Participant #9: ... the average number, the fact that there's a max, and not wanting to... In fact, when I was doing it, I thought, "If I'm doing an experiment, I could make sure that my prior thinks that there's no difference between the two conditions, or even the opposite. To say the prior should be even more conservative for the condition to make sure that the evidence is sufficient."

Participant #9: But I decided just to go with, "Well, let's assume that doesn't have an effect and let the likelihood contribution actually help us to distinguish them." So yeah, I think the same thought process for both. The setting the intercept and the mean difference.

Interviewer #1: Mm-hmm (affirmative). I guess my question is, what kind of information do you look for in the second visualization?

Participant #9: In the second visualization?

Interviewer #1: Yeah. They chose the normal, the centered at zero and a scale of 0.61.

Participant #9: This becomes tricky, right? Because this is setting...the previous studies speak more to the intercept. So on average, how people will respond. I think maybe the main thing here is, was looking to make sure that there was no difference between the two conditions.

Participant #9: So something like a standard normal seemed to be-

Interviewer #1: I see.

Participant #9: ... fine. I guess the thing that I maybe would want... Was there an option to have this be centered around zero?

Interviewer #1: No, because this is the prior predictive density, so if you drag around along the X-axis, then you would realize... So if you drag it towards one side and it'll show that it predicts more for one condition to be greater than the other one.

Participant #9: Right. Okay, so I'm thinking back to what I actually did. Yes, I think given that it couldn't be centered around a null effect, I think I just tried to find the condition that had them be essentially equal. So yeah.

Interviewer #1: I don't think I understood that. What do you mean by null effect?

Participant #9: The effect of the condition would be zero. So I just had to find that the difference between the two, the distribution was the same, right? And trying to make sure that this... It's not like this, right?

Interviewer #1: Mm-hmm (affirmative).

Participant #9: I want to make sure that something like this... That's probably why I didn't change it much.

Interviewer #1: I see. I mean, I guess what I'm having difficulty is what kind of visualization were you thinking when you say a visualization of the null effect?

Participant #9: So this is for the... This density is all set over... There's nothing that allows me to have a really peak distribution for both of them set around zero, so that this mean difference parameter could be zero. There's just a low probability of it being zero here. That's probably a scale thing that I'm not getting.

Interviewer #1: I see. Okay.

Participant #9: Yeah, I couldn't set this at a... Because if you think about it, it's like a standard regression model. Then my distribution, my prior for some estimate is going to be centered around zero, right? There has to be evidence that it's different from zero. But here, I guess it was just that the two distributions were essentially the same.

Interviewer #1: Mm-hmm (affirmative). Okay. Could you talk about did you face any challenges while interacting with this interface or did you wish that you had some other information that was not presented on this interface to you?

Participant #9: I guess looking back, maybe the thing that I was missing the most was if it's on a log scale, that always kind of throws me for a loop. So setting a prior on a standard scale. That may be a little bit easier to interpret. It may have been helpful.

Participant #9: It would have been nice to be able to type in exact values instead of just dragging them around as well. I mean, I like exploring this, but it might have been nice to be able to type in what the mean and standard deviation are.

Interviewer #1: Mm-hmm (affirmative). Okay. When you mean the standard scale versus the log scale, do you mean if instead of having the priors presented on the log scale, we presented it on the standard scale?

Participant #9: And maybe I'm just not thinking about this the right way. But yeah, if it was on the standard scale, then maybe it would be easier to translate this information, right?

Interviewer #1: Mm-hmm (affirmative).

Participant #9: About the average number of pumps and the possible number of pumps.

Interviewer #1: Okay. So basically, the lambda should have... Because in the equation, if we say we are using a log transformation on lambda... Yeah, I think.. Okay, I see what you're saying. I think I do.

Interviewer #1: Cool. Let's move on to the next page.

Participant #9: Okay.

Interviewer #1: Again, I kind of mentioned this in the beginning, but this is the second page of three pages.

Participant #9: Okay.

Interviewer #1: So this is going to show you three different visualizations. You've seen one of them before already. The other two would be new ones.

Interviewer #1: Again, looking at this first situation, this is just representing the density of the parameter. The previous one, where you were showing the prior predictive, that is like we fit the model using the priors as the data generating process. And show you densities from a hypothetical experiment. Whereas this is just the density of the parameter on its scale. Looking at this visualization. Could you talk a little bit about how you would use this information to set your prior?

Interviewer #1: (silence)

Interviewer #1: I see that you're thinking some things, could you-

Participant #9: Yeah. No, I'm just trying to think about what I'm looking at here.

Interviewer #1: This visualization is literally the visualization of... The density of. Let's say you-

Participant #9: Just the-

Interviewer #1: ... put it to a corner and a normal at location three and scale 0.2, what that would look like.

Participant #9: Then the X-axis, this would be just the parameter value. This would be for-

Interviewer #1: Just the Y-axis and density.

Participant #9: ... the intercepts. Yeah, okay.

Interviewer #1: And then we have the model and you have to interpret it in terms of the likelihood and yeah.

Participant #9: To be honest, giving feedback for this is... In my training, the idea of a prior predictive distribution is actually not something that I was ever trained with. That's something that's relatively new to me.

Participant #9: So when we set priors, we do it more like this. I'm just trying to think about what the scale represents. So this is just the intercept and the model for the number of pumps, but this is on the log scale again?

Interviewer #1: Yes.

Participant #9: So yeah, I'd say that this is harder to think through. The prior predictive distribution gives me kind of the expected value, so I can more directly use the information that was given about previous studies. Because I can see that I should be getting... Maybe missing the log scale stuff, but I'm getting results that are... My prior isn't producing results that are inconsistent with what I've seen before.

Participant #9: Here, I have to think about what... I probably need more domain expertise to set a prior directly in this way. I have to think about what that parameter value should be explicitly.

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

Participant #9: Yeah, this is certainly harder to do.

Interviewer #1: ... Now, could you briefly talk about what strategy would you use using this visualization? You don't have to actually implement the strategy, but overall like high level, how would you go about it?

Participant #9: Sure. I think in the absence of any... Direct way to translate the domain expertise. I'd probably just make this as diffuse as I could.

Interviewer #1: Mm-hmm (affirmative).

Participant #9: Something like this.

Interviewer #1: Where would you place your mean?

Participant #9: Yeah, I'm not sure. I'm not sure in terms of that model translating what the meta-analysis said to this. Yeah, I can't do that calculation in my head. I'm not sure what that would be.

Interviewer #1: I mean, what calculation would you try to do, even if you're not doing the calculation?

Participant #9: Yeah. I just need to think about in terms of this is on the log scale, then how do I take the expectation from the meta-analysis and turn this into making sure that the prior is over reasonable values. I'm not sure what that would be. I guess if it was just from zero to six in this case, then I would make it something like this [Normal(3, 1)].

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

Participant #9: Just kind of middle of the road. Fairly diffuse.

Interviewer #1: Mm-hmm (affirmative). Okay, cool. Do you have a preference for normal or student's t or something?

Participant #9: Yeah, I probably have more of a preference for normal.

Interviewer #1: Okay. Why is that?

Participant #9: Probably just easier to intuit what the parameter values and the prior represent. Even though I can see the visualization...familiarity, I guess.

Interviewer #1: Mm-hmm (affirmative). Cool. Let's move on to the next one.

Participant #9: Okay.

Interviewer #1: This is the same. This is the prior density, but we're using the transformation. So we're transforming it to the scale of...the exponential scale. So you're not interpreting it on the log scale anymore. It's the same density, but transformed.

Participant #9: Okay.

Interviewer #1: So because if you transform a normal density, which is on the log scale, and you transform it into the natural scale, then you get a log normal density. So this is roughly what that is.

Participant #9: Okay.

Interviewer #1: Again, if you look at this information and if I were to ask you to set a prior, how would you use this information if at all? Or would you just stick to the prior that you chose in the prior predictive density?

Participant #9: Read this real quick.

Participant #9: Okay. So this is on the scale of pumps.

Interviewer #1: Right.

Participant #9: So I think this would be easier to set, right? Because now I can directly use that information I had before.

Interviewer #1: Mm-hmm (affirmative).

Participant #9: Whatever the meta-analysis was... This is probably what I was trying to do with the prior predictive check, right? I'm not sure which of these I would use. But yeah, setting so it can't be over whatever the max was, 120 something?

Interviewer #1: 128, yeah.

Participant #9: 128? What was it, the range was like 24 to 60 or something?

Interviewer #1: No, 24 to 45.

Participant #9: 24 to 45.

Interviewer #1: That was the expectation of the mean.

Participant #9: Yeah.

Participant #9: So yeah, something like that. Yeah. This is probably what I was trying to do before, then I didn't think about being on the log scale.

Interviewer #1: Yeah. Cool. That makes sense.

Participant #9: Yeah, I don't know. I think that's probably the most straightforward way to think about it. Translating what I had before directly into setting the prior.

Interviewer #1: So I guess looking at this, could you contrast the information presented here to the previous one, which was on the log scale?

Participant #9: Here, I don't have to think through the translation of what the scale I'm setting this prior over, what the range I'm setting this prior over is on the log scale. It's directly in terms of pumps, which is how the information was given to us.

Interviewer #1: Mm-hmm (affirmative). I'm just curious. Have you ever used visualizations in your own experience for setting priors? Especially when you're using nonlinear models like this one, how do you go about choosing priors?

Participant #9: Not a great deal of visualizations. Again, I like the idea of a prior prediction check, because it's kind of new to me. A lot of times, just kind of standard conjugate priors and priors that have been set. I often work with studies that the models had been run previously. So kind of use rules of thumb. Kind of use standards.

Interviewer #1: Mm-hmm (affirmative). Have you ever had to do these sorts of transformations or stuff like that? Or setting a prior.

Participant #9: Not a great deal, no.

Interviewer #1: Mm-hmm (affirmative). What kind of models do you usually fit?

Participant #9: I do a lot of hierarchical multinomial logit models.

Interviewer #1: Okay, cool. That's interesting.

Participant #9: It's the bread and bread butter of marketing.

Interviewer #1: I see.

Participant #9: Yeah.

Interviewer #1: I'm curious, what kind of conjugate priors do you use for a hierarchical... I mean, I'm not aware of that, so I'm curious to know what it would be.

Participant #9: So yeah, this is kind of showing my... Training isn't that old, but my trainer is kind of dated. This is prior to any sort of use of Hamiltonian Monte Carlo. So if you're using a random mock Metropolis for a hierarchical model, then some amount of conjugacy matters.

Participant #9: We typically have a Gibbs within Metropolis algorithm, and you have normal priors, hyper priors, and inverse Wishart priors on the variance.

Interviewer #1: Yeah.

Participant #9: Yeah. So there's-

Interviewer #1: I remember.

Participant #9: Yeah, there's standard one and it's... I know it's not good practice and I know there's some interesting work in terms of... I mean, part of the challenge for us is it's not even super clear what is the best practice for to create.

Participant #9: What do you visualize when you're trying to visualize a prior predictive check? Because we're dealing with a whole bunch of discrete data. We're looking at choices. These are models of people's choice behavior and try to estimate their preferences. So I like the idea of this, but I'm probably super awkward at it because I haven't done a lot of nondiscrete choice models.

Interviewer #1: Mm-hmm (affirmative). I haven't seen any resource for prior predictive checks for those kinds of models, I think.

Interviewer #1: There's this statistician called Richard McElreath. I don't know if you've come across him.

Participant #9: Mm-hmm (affirmative). Yeah. Absolutely.

Interviewer #1: His book is called Statistical Rethinking. His first book has some stuff, but I think his newer edition, which is coming out this year or next year will have more of prior predictive checks.

Participant #9: Yeah. I'm may be teaching from his book for the first time in a pre-PhD class this semester.

Interviewer #1: Cool.

Participant #9: Yeah, so [crosstalk] first edition.

Interviewer #1: Yeah, I love his book.

Participant #9: Yeah, it's really great. It's fantastic. Maybe I should do this again after I go through the class. I'm sure I'd be better at it.

Interviewer #1: Cool. Sorry for going off track.

Participant #9: No, that's fine. That's good.

Interviewer #1: I guess we can move on to the next page.

Participant #9: Okay.

Interviewer #1: Yeah, because you've seen this one before.

Participant #9: Yeah.

Interviewer #1: So if you scroll down and click on the next button.

Participant #9: It's just go to the next one?

Interviewer #1: Yeah. This is the last page, and then this is exactly the same as your previous one except we are going to be working on with the mean difference parameter. Again, if you look at the visualization and talk a little bit about how you would use it, just takes a while to load. It's because we're loading all the data beforehand.

Participant #9: Nice. This is all hosted on GitHub?

Interviewer #1: Yeah.

Participant #9: That's awesome.

Interviewer #1: Again, this is the density. If you're setting a prior on the normal zero one, this is what kind of look like.

Participant #9: This is on the-

Interviewer #1: This is 0.6. It's on the log scale.

Participant #9: This is on the log scale? Yeah. So if I think about the difference between setting a prior using the prior probability density versus using the prior predictive density, and the prior predictive density, I probably should have compared more between the prior I was setting for the intercept and the different... mean,difference parameter. Because I think the mean difference parameter, whatever prior I set should have been very similar, right?

Participant #9: And maybe that's what I did. I would make sure they're similar here. This is I think what I was talking about, right? It's I want to make sure that this is centered around zero because I should have a prior belief that there is no difference, and then the data has to inform whether there's a difference.

Interviewer #1: Mm-hmm (affirmative). What kind of a prior would you choose in this case?

Participant #9: Something centered around zero for sure. I'm not sure how diffuse there was information in the setup about some variants in previous studies, but I'm not sure if it applies directly here.

Interviewer #1: Yeah, we didn't provide any information regarding the, at least, what to expect regarding the two conditions.

Participant #9: Sure. So I probably would set something pretty diffused centered around zero here [Normal(0, 1)]. I'm not sure what the size of the effect could be and I don't want to set anything dogmatic that would not allow for a larger effect.

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

Participant #9: Something like that.

Interviewer #1: Yeah, I guess that makes sense. Cool. I guess we're going to move on to the next one.

Participant #9: Okay.

Interviewer #1: So this is the density, but it's on the response scale. So I guess the differences on the axis where a linear additive term on the log scale is essentially affecting the data in a multiplicative manner. So basically, this is showing you how much probability are you giving for... What is your probability for expecting an effect? Which is twice as large in the test condition versus the controlled condition.

Participant #9: Yeah. So instead of being set around zero, it's just set around one times?

Interviewer #1: Right.

Participant #9: Yeah, I'm not sure if this is easier to think through. Again, because I'm not sure I know how large the effect should be. So again, I'll just keep it very diffusely set around one times. Because I mean, I don't know what to expect.

Interviewer #1: Mm-hmm (affirmative). So you would almost want a domain expert to tell you what kinds of effect sizes are reasonable in studies like these?

Participant #9: Sure. Yeah, absolutely. At least that would allow me to set a more informative prior.

Interviewer #1: Mm-hmm (affirmative). Would you still center it at one or would you center it at something else?

Participant #9: I'd center it at one still. But I think this one is maybe more intuitive to me of these two. It kind of feels like an eye exam, actually. This one versus this one.

Interviewer #1: I guess I'm curious to know how do you think of the distribution in the previous one, how it affects the model?

Participant #9: For this one?

Interviewer #1: Yeah.

Participant #9: Setting a diffuse prior around zero, saying that we expect there to not be an effect. If it's diffused, then there could be an effect, but we're not sure what it is.

Interviewer #1: But how large an effect are you permitting-

Participant #9: I'm not sure.

Interviewer #1: ... based on this?

Participant #9: I'm not sure. Right? I'm trying to set it to be as diffused as possible because I'm not sure. These ones becomes very low probability for it to be a large effect, right? It's got to be somewhere around zero. That's about as diffuse as I can make it. Yeah.

Participant #9: (silence)

Interviewer #1: Hello?

Participant #9: Hello?

Interviewer #1: Hi. Sorry, I'm having some network issues.

Participant #9: That's okay.

Interviewer #1: You'd imagine a university to have good network, but... Cool. I was wondering if you were not constrained by this widget thing, would you almost had a uniform prior?

Participant #9: I would probably not feel comfortable setting a uniform prior, but I might want to make it more diffuse than this.

Interviewer #1: Mm-hmm (affirmative). Okay. Cool.

Participant #9: If I were being thorough, then yeah, I would definitely want to be able to... Yeah, see what the resulting prior predictive distribution would be.

Interviewer #1: Mm-hmm (affirmative). Cool. I still feel like your expectation with the prior predictive distribution versus what it's actually showing you, are not aligned.

Participant #9: Okay.

Interviewer #1: I mean, I guess this is a chance for us to improve how we're showing the prior predictive distribution, but... Could you talk about the other visualizations and contrast it to the information presented here? And what could have made it easier for you, looking at the prior predictive distribution to make your decision?

Participant #9: Maybe I missed the scale thing. I think scales was maybe the biggest disconnect. Because the prior predictive distribution was giving us the response on a log scale, right?

Interviewer #1: No, it was on the natural scale.

Participant #9: It was on the natural scale?

Interviewer #1: Yeah.

Participant #9: Okay. Yeah, so then I can use the information directly, right? That was given. And maybe that's what I did, right? Just make sure that it's within the expected range in terms of the number of pumps that aren't possible, given the constraints of the experiment.

Interviewer #1: Mm-hmm (affirmative). I guess one thing that was not super clear is how we calculated the prior predictive distribution, and it sort of... It's a joint probability distribution, right? So what you said for the intercept parameter will affect your mean difference parameter.

Interviewer #1: I mean, that is something that was very difficult to calculate. So what we did was we fixed the... When we're showing you the mean difference parameter, we fixed the intercept parameter at a fixed value and then we showed you.

Participant #9: Sure.

Interviewer #1: Yeah, I think that may have also caused some confusion.

Participant #9: Yeah.

Interviewer #1: Okay.

Participant #9: Yeah, maybe. Yeah.

Interviewer #1: Yeah, I think that was a hard visualization too for us to create.

Participant #9: That's cool. That's a cool one. I'm sorry if I missed something about the scales. Coming back to it now, that's what I'm most confused about.

Interviewer #1: No, that's perfectly fine. I think that serves us... That's actually valuable information for us because it's something, I mean, that can be missed that was not presented clearly enough.

Participant #9: Or I just don't remember, having read it before. So maybe I did it right the first time. I don't know.

Interviewer #1: It's also been like two weeks, three weeks.

Participant #9: Yeah, it's been a little while.

Interviewer #1: Cool. I guess I have a few final questions for you.

Participant #9: Sure.

Interviewer #1: Is there anything else besides what you mentioned regarding conjugate priors in your work, what information do you use when selecting priors?

Participant #9: I wish I had more general experience with setting priors. It's really kind of tailored within this specific niche of choice models and logit models.

Interviewer #1: Mm-hmm (affirmative). I guess if you could talk about that as well, that'd be helpful for us.

Participant #9: Yeah, sure. I mean so again, that literature in marketing in particular, comes from a still sensitivity to conjugacy. I guess I've seen some work around at least the beginnings of what would become the prior predictive distribution, in terms of setting an inverse Wishart prior and how it can be unintentionally informative. Which is probably why we would stick with certain standards of terms of like a good rule of thumb for setting that prior or these specific values.

Participant #9: But setting values over various parameters is kind of tricky to begin with. So you kind of lean on like, "Well, this is the way it's been done. So that's the way that we're going to do it."

Participant #9: Again, in the absence of really good visualizations for a prior predictive check, it's hard to kind of seriously interrogate the way that we're setting priors.

Participant #9: Jim Savage, who does more like Bayesian econometrics, but that kind of abuts the logit model stuff. He's done some work... He presented at StanCon last week a little bit about this.

Interviewer #1: Cool.

Participant #9: But there's like no consensus in terms of what that prior distribution would be.

Participant #9: To be honest, in my field, there's not a lot of talk about the fact to even make these, right? You kind of just still use... There's not a lot of sensitivity to being thoughtful about setting priors generally.

Participant #9: So coming out of that, I'm interested in being more sensitive and using kind of modern tools to be able to figure that out. But I don't have a lot of intuition about setting them generally.