Interviewer #1: Reiterate that the purpose is not to judge your ability to set priors, but for us to understand what sorts of interfaces might help people better set priors in general.

Participant #3: Mm-hmm (affirmative).

Interviewer #1: If something doesn't make sense, or something doesn't seem useful, we really want to know how we can improve it, in the feature designs. If you think that something is useful, it is important for us to know that as well.

Participant #3: Yeah.

Interviewer #1: Cool. This document consists of three pages, the first page is exactly what you saw in the survey, at the second page of the survey, you saw the prior probability density visualization.

Participant #3: Mm-hmm (affirmative).

Interviewer #1: You set the prior as 3.5, and the standard deviation as 0.7. You chose a normal distribution. Looking at this interface, and the prior that you chose, could you walk me through what you did, and how you decided to choose this prior?

Participant #3: Yeah, the first thing I had to do here was, I had to remember what the story was. I clicked, "Show description," I believe. Yeah, and recalled that you have this [inaudible] model here, with this sort of linear model that you need to convert, on a scale of zero to infinity, by using the log, right?

Interviewer #1: Mm-hmm (affirmative).

Participant #3: This is the ... I forget the name of this transform, inverse [logit 00:01:39], is that what it is?

Interviewer #1: Inverse log.

Participant #3: Inverse log, yes. Because the intercept term here was on this scale that ... I'm not that fluent in jumping between a logarithmic scale and a linear scale, I went over to my R window and simply typed in the exponential function of zero, or one, or whatever it was, just to try to get my bearings around how many pumps are reasonable. What does that mean for alpha, so I wanted alpha to be able to represent this average number of pumps. You said, from past research, 24 to 44, I wanted it to be wider, to include other elements. I tried to get into the ballpark, and I assume that if you take the exponential function of 3.5, you get a number of 33.

Interviewer #1: Mm-hmm (affirmative).

Participant #3: That was right in between, and close enough, in between 24 and 44, as the expectation of the average number of pumps. Then, I probably thought to myself, "I need to go two standard deviations on either side, and see what effect that means." If I recall correctly, I took the exponential function of 4.9, and the exponential function of 3.5 minus 1.4, it's 2.1, and that got me to eight. I said, "Okay, this range of expected pumps, from 8 to 134, I'm pretty happy that's a wide range, that could include or capture any of the data that we see."

Interviewer #1: I see, so you basically took a more bottom-up approach, where you decided that the prior would be reasonable, based on the information given, and then you verified if it matches the experiment design?

Participant #3: Yes. To me, it's much easier to think on the outcome scale.

Interviewer #1: That's interesting. That's an interesting approach.

Participant #3: Then, I had to choose between normal or Student-t, and for the intercept, I said, "I'm pretty confident that this number is going to be between 0 and 128, I don't need to accommodate fat tails," so I picked normal.

Interviewer #1: Cool, okay. This is great. Then, let's move ... And for the mean difference?

Participant #3: I knew, for the mean difference, I wanted to be centered on zero. If we go to the visualization, I knew zero was appropriate, and then I wanted to indicate, "Well, how fat do I want my tails?" My guess is, I took exponential function of, just to get an idea, and I kept going back to my little R calculator here, but ... Moving from 3.5 to 3.7, and the exponential function got us seven units of difference.

Interviewer #1: Mm-hmm (affirmative).

Participant #3: To me, if we look ... Let me remember the problem. "Risk introduced with each pump, the more a player pumps up the balloon, the higher their payoff, the risk is introduced around the point of explosion, with average, and mean being explosion point at 64 pumps. Repeat this 60 times, psychology research ... " I forget how the condition was applied.

Interviewer #1: Oh, participants were either in the controlled condition, which means they were in the constrained position, or in the expansive condition, which is the test condition, so they were sitting with their arms wider apart.

Participant #3: Ah, yes, yes. Just, my prior experience with power poses or whatever this is, I figured that the effect is not going to be that large. I want to be close to zero, but I wanted a T distribution because maybe I'm wrong, here. Maybe the effect can be large, so I wanted fatter tails. I knew it should be centered on zero, and then I just ... Where is the zero?

Interviewer #1: That's fine, it's in the middle. Yeah.

Participant #3: Yeah, there we go. I just went up and down like this and probably said, "What makes sense for a small effect size?" Oh, and I went to an extreme point, here. That's because I felt the effect of this will actually be smaller than this distribution suggests, based on the prior.

Interviewer #1: I see, so you wanted it to be even smaller? You wanted the prior to be even narrower?

Participant #3: I would have gotten down even tighter. I would have stayed with a T.

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

Participant #3: That's my prior distributions, based on these poses, I don't think they really have an effect ... I would've-

Interviewer #1: [crosstalk 00:08:20]. Oh, sorry.

Participant #3: It's also a function of how much data you have. If you have a lot of data, then I don't need it so much either, I know the data will overwhelm the prior. If there's not a lot of data, then ... I felt my prior should be tight.

Interviewer #1: Mm-hmm (affirmative). My follow-up question is, how did you interpret the tails of the distribution? How do you think that will affect the difference between the two conditions?

Participant #3: The tails, why I chose the Student-t as opposed to a normal, here?

Interviewer #1: No, more how ... In general, any distribution, regardless or certainty or normal, how did ... How are you interpreting the probability for this to affect the model, essentially?

Participant #3: I always convert to the outcome scale. I do some sort of sampling of variables that I think can come from these two distributions, and I make sure that it admits possibilities on the outcome scale that I think are reasonable.

Interviewer #1: I see. You just sample from the mean, and the mean difference parameter, and the intersect parameter, and then you convert that into the response scale.

Participant #3: Yeah.

Interviewer #1: Okay, I see. Interesting. Okay, we can proceed to the next page.

Participant #3: Okay. I think about this stuff a lot, and [inaudible] showing up, that paper, with the quantile dot plots?

Interviewer #1: I see.

Participant #3: I think it's effective. Often, I think that these density functions are hard for us to interpret accurately.

Interviewer #1: Yeah. Again, we can ignore the first plot, and just focus on the second one, because the first one we already looked at, just now.

Participant #3: Okay.

Interviewer #1: Just to give you a quick blurb about this, this is the same probability density for the intercept, but we are translating it into the response scale, so-

Participant #3: Ah, this is what I was hoping for. Cool. Yeah, I love this.

Interviewer #1: My question is, looking at this visualization, do you think that the information presented in this visualization would affect your choice of priors, and how would you use this information, if you're using it at all?

Participant #3: Yeah, this would absolutely be how I want to choose my priors, and I don't even care about the parameters themselves. I really just want to distribute my prior probability on this outcome scale.

Interviewer #1: Mm-hmm (affirmative).

Participant #3: Now, I want to center it around 33, somehow ... Have things greater than 128 not be common, lower than zero being possible. Yeah, this is great. Something like this is cool, I know that my ability to interpret this density function is a little limited, but I'd love to see samples drawn from the function.

Interviewer #1: You mentioned that you would try to center it around 33, how would you do that? How do you go about doing that? What do you look for, in the visualization?

Participant #3: Initially, I was looking for the mode to be around 33, but as you bring it up, I should technically have tried to distribute half the area below 33, and half above 33. As long as the hill is somewhere in the 30s, I'm okay. My prior knowledge is not accurate to a tenth of a decimal point.

Interviewer #1: Yeah. We don't expect-

Participant #3: I'm fairly comfortable with-

Interviewer #1: Yeah, we don't expect it to be ... The resolution on this scale is too much, I think, for a prior, because of the transformation and everything.

Participant #3: Yeah. It's great. I love it.

Interviewer #1: Would you change your prior in any way, in this case?

Participant #3: I don't remember what my old prior was ...

Interviewer #1: It was 3.5, and 0.7 as the standard deviation.

Participant #3: 3.5, and 0.7. It was here ... Yeah, I felt like moving it this way [changes the location of the prior (on the intercept) towards 4].

Interviewer #1: Why is that?

Participant #3: Because at 3.5, 0.7 ... Let me find it again. It puts too much mass on below 33, on small outcomes, to me. I want more area to the right of 33, because I don't really know what I'm talking about with these pumps, so I would like to stretch it out more, like that. I'd like to go this way.

Interviewer #1: I see. Cool. What is the main contrast that you see with this information, and the previous visualization?

Participant #3: The main contrast is that I'm on a scale I can understand, now. As far as that exponential function scale, transformation, my brain doesn't work like that.

Interviewer #1: Mm-hmm (affirmative), yeah. Okay, let's move on to the third visualization, which is the prior predictive probability density. I think this is what you were talking about, you mentioned this in the survey. You had mentioned that you would rather just draw samples from the prior predictive distribution, instead of looking at this. Yeah, could you tell me what you think of this information, and how it would affect your choice of priors, compared to the other two visualizations that you saw?

Participant #3: With this, prior predictive density for 20 hypothetical experiments ... This one's a little tougher, I don't like the transformation to some sort of kernel density function estimate.

Interviewer #1: Mm-hmm (affirmative).

Participant #3: I don't know how much data is in the hypothetical experiment.

Interviewer #1: I see, it's 40 participants in each condition.

Participant #3: 40 participants, so I'd rather actually see a quantile dot plot drawn from prior, and maybe see a bunch of them.

Interviewer #1: Could you elaborate on that?

Participant #3: These are prior predictives ... When I look at this information here, I know something is wrong. One is suggesting that more than 200 pumps ... This is not the average, or is it? For total number of pumps that participant ... No, it can't be that.

Interviewer #1: No, it's just showing that one participant may have pumped ... In one trial, they may have pumped 200 times, because of how we model the data, if you look at the parametrization of the model, we don't have an upper bound.

Participant #3: Mm-hmm (affirmative).

Interviewer #1: We haven't set an upper bound, and also, the Student-t distribution has those fat tails, which ...

Participant #3: Yeah, and this is the prior on the intercept, or-

Interviewer #1: Intercept.

Participant #3: The prior on the intercept, yeah. I'm finding, in this upper region, I'm getting, at least, on a scale that makes more sense to me, but even ... I'd like to see, if it's 40 people doing the experiment ... I would like to see a histogram of the number of pumps, for the 40 people from a prior.

Interviewer #1: I see, and not something like a kernel density estimation.

Participant #3: Yeah, this spaghetti plot ... I think you lose the granularity of the experiment itself.

Interviewer #1: Mm-hmm (affirmative).

Participant #3: Oh, and I see these colors here, constrictive versus expansive.

Interviewer #1: Yeah, in this case, it doesn't make a lot of difference, because the prior there has been a separate prior chosen for the data, and ... When you're changing the intercept, those are overlapping, almost. If you use the dropdown and look at the normal, do you see any difference?

Participant #3: Oh, certainly. The normal plots are nicer, in that I like the scale better. The T plots, with the fat tails, didn't get as narrow as this. I like that one better. I still find the spaghetti kernels challenging.

Interviewer #1: Mm-hmm (affirmative), I see. You mentioned something interesting about dot plots, can you tell me how you would use dot plots in this case?

Participant #3: For a prior predictive, show me 20 dot plots of the hypothetical experiments, and it could be faceted. I think, maybe having them all on one plot is too confusing.

Interviewer #1: Mm-hmm (affirmative), so facet, by each hypothetical experiment? Or faceted by condition?

Participant #3: Faceted by experiment.

Interviewer #1: Mm-hmm (affirmative), I see, so it will be 20 plots. Okay. Yeah, so-

Participant #3: Or even two plots, and then you can hit refresh and get a new one each time, and cycle through them.

Interviewer #1: Finishing up with this, would you change your prior in any way, given this visualization, or would you stick to ...

Participant #3: The prior predictive? Would I change my prior ... As I went up here, I said that I don't have enough mass to the left of 30, so that's not good. Yeah, my instinct is that I shy away from these, because it's hard to interpret. I like this one [Normal(4, 0.2)], Just because it seems parsimonious, but when I study the details, this doesn't really capture my prior information, so now I'm forced to go in the sort of zone where I think I was. No, I was in the Student-t zone. Yeah, the Student-t zone, I would want to truncate it. Truncate my prior, that is. Something like this, is not terrible [Normal(3.9 0.6)]. I don't know, what was I at before, 3.7 and 0.5?

Interviewer #1: Your original was 3.5 and 0.7.

Participant #3: 3.5 and 0.7 ... For the Student's-t.

Interviewer #1: No, you had a normal.

Participant #3: Oh, I had a normal for the intercept, right. You're right. Yeah, I almost want to admit higher numbers. I'm not dissatisfied with my original prior, but the scale change makes it a challenge, with this visualization.

Interviewer #1: That the scale that constantly keeps changing?

Participant #3: Yeah, the fact that the X axis scale is changing is problematic.

Interviewer #1: I see. Mm-hmm (affirmative), okay. Again, just for recap, could you contrast this information, the information presented here, versus the other two visualizations?

Participant #3: This plot, for me, is adding noise that I have trouble interpreting. It does make me want to truncate this distribution, seeing all of these little humps here to the right of 128. That's not terrible. I feel that this plot, up here ... I get all the information that is shown down here without so much noise, that my brain is struggling with.

Interviewer #1: Mm-hmm (affirmative), okay, so let's move on to the next page. This is the last page of this and, again, we're going to go through the same thing that we did for the previous one, but we're focusing on the main difference parameter.

Participant #3: Okay.

Interviewer #1: This takes a while to load, but if you scroll down to the second visualization ... Okay, this is, again, the prior parameter probability density, and your response scale. There are a few changes from the other one. Again, when you look at this, what are you ... What information are you looking at, and do you think this is going to affect your choice of priors in any way?

Participant #3: Now that I see this, this is how I set my priors originally, that I thought 20% in either direction seemed reasonable, and when I went to ... I know it's up here somewhere. Right there, yeah, I wanted to go tighter, because I don't think it's realistic to think that this power pose is going to cut pumps in half, or double them. For me, from ... I'm a business guy, so I'm more practical than experiments, where I'm trying to get a P-value type of thing. Yeah, I want this tighter, but I love this scale. This is the scale that I wanted, initially. This multiplier effect.

Interviewer #1: I see, so when you're looking at it on the scale of the log scale, you were trying to do the back calculation of what would be the multiplier effect?

Participant #3: Yeah, and for me, almost a uniform distribution from 80% to ... From 0.8 to 1.2 would have been cool. This, it just doesn't get tight enough for me, but it's okay. The visual is very good.

Interviewer #1: You mentioned that you'd be fine with a uniform distribution, with that very narrow, 0.8 to 1.2 or something.

Participant #3: Yeah.

Interviewer #1: Could you expand on that a little bit? Why would you use a flat prior?

Participant #3: Just because I think that the data will quickly inform us, as to the magnitude of the effect size, and I don't know of a great prior for representing my knowledge that I think that the effect size is going to be somewhere ... I'm very confident that the effect size is between negative 20% and positive 20%, without admitting a lot of tail probability. I find that a lot of calculation can get thrown off by having extreme tail possibilities, that I just want to ignore. I downright call them infeasible. I guess that I'm willing to substitute a flat prior over an interval I'm 100% confident in, than a normal or T prior that has these fat tails, admitting many, many things that I'm not confident in.

Interviewer #1: Mm-hmm (affirmative), I see, so you would rather choose a flat prior, over a very, very thin normal distribution.

Participant #3: Yes.

Interviewer #1: That's interesting.

Participant #3: I think the spikes here, it just doesn't feel that right to me. You know what a Beta(2,2)distribution looks like?

Interviewer #1: I think so, it's it almost a flat at the top, and very steep tail?

Participant #3: Yes, up and down U. That looks a lot like a Beta(2,2) shape, going from 0.8 to 1.2.

Interviewer #1: Okay. Let's move on to the last prior predictive probability distribution.

Participant #3: Okay.

Interviewer #1: Yeah, again ...

Participant #3: I like the separation, but here ... I know my mean should be zero. I want to say that there's no difference. Once we lock in no difference, the fact that the X axis scale keeps changing on me makes it hard for me to take advantage of this visual. I'd almost want the ... I do want the prior predictive on the difference.

Interviewer #1: What do you mean?

Participant #3: I'd rather have the prior predictive on the beta parameter? Yeah. The beta?

Interviewer #1: I don't see what you're ... I don't think I understand what you're trying to say, could you elaborate on that?

Participant #3: Yeah, let me ... Just trying to formulate my thoughts, here. I know the difference between, or the causal impact of the condition ... I know I want that centered at zero, but I also want to know ... In terms of number of pumps, or in terms of this multiplier effect like we had here, on the X axis scale, I'd like to know the impact, either on this multiplier effect scale, am I admitting large deviations of pumps between the two conditions, am I centered around small deviations of pumps between the two conditions ... I just don't get that from this diagram, what kind of change am I saying is plausible with this prior.

Interviewer #1: I see, that's interesting. Okay.

Participant #3: I know changes of 50 pumps could not be plausible, in this experiment, and I don't know that I ruled that out. I know changes of five pumps was very plausible, and I can't see that from this diagram.

Interviewer #1: I see. Mm-hmm (affirmative), that's interesting. Okay, cool. The few other questions I had was, what other information do you think would have been helpful for you, when you were setting this prior?

Participant #3: I'd like all of the outcomes ... When I'm setting priors, I really need to go quickly from parameter to outcome scales that I'm interested in, and I guess there are several ways of expressing the outcome scales, or what outcomes I'm interested in. One is just the generic, what's the average number of pumps which is impacted by the intercept, and the other thing I'm really interested in is, what is the deviation, probably both in relative terms, like this X axis you have here with the multiplier effect, as well as in absolute terms. What does that mean for the number of pumps, if we had someone change their arm position?

Interviewer #1: Mm-hmm (affirmative), I see, so in the second visualization, if instead of showing twice, it shows you that the mean for the intercept was 35, then the mean for the test condition can be either up to 70, or as low as 17.5, or something like that.

Participant #3: Exactly. Show me what scale of possibilities we're thinking about.

Interviewer #1: Okay, and-

Participant #3: The other thing I'd mention was, I would look for alternatives to the kernel density plots, potentially.

Interviewer #1: Interesting. You mentioned alternatives, using dot plots, or histograms?

Participant #3: Yeah, exactly.

Interviewer #1: Is there anything else?

Participant #3: Is there anything else ... I can imagine, where I could see something, here ... That one. These kinds of plots (showing forest plots of point estimates and credible intervals, http://causact.updog.co/16-MultilevelModelling\_files/figure-html/alphaBetaEstPP-1.png), I like, but maybe on the outcome scale, instead of ... This is on the parameter scale, but I can imagine 20 hypothetical experiments on the horizontal axis, and then you can have two bars, one for ...

Interviewer #1: [crosstalk 00:39:17]-

Participant #3: The mean pumps without condition, and another for the mean pumps with the condition.

Interviewer #1: I see, so you would almost parametrize your model differently, in this case. Instead of having a beta, you would do alpha one and alpha two, maybe.

Participant #3: Instead of the beta, I would have ...

Interviewer #1: Alpha control and alpha test?

Participant #3: Alpha control and alpha test, and then it would be a prior predictive, so you would have, maybe, ten of them, and show the feasible ranges for the number of pumps in each condition.

Interviewer #1: That's very cool, that's very interesting. Okay, I guess, wrapping it up-

Participant #3: [crosstalk 00:40:10], it's [inaudible] thing, but I like those graphs, because I think I've abandoned using kernel densities, and just use these bars, because they're more compact. Since this is 20 hypothetical experiments here, 20 bars is a lot easier to digest than 20 spaghetti kernel things.

Interviewer #1: Mm-hmm (affirmative). Okay, so my final question is, can you think of other instances in your past, while you were choosing priors for models, what information did you consider, did you face any challenges while choosing priors, and-

Participant #3: Oh, it's always a challenge, and as soon as you scale your model, getting to the outcome scale ... Because that's really where your prior information lives, for the most part, is on the outcome scale. Certainly, from converting any sort of hierarchical model into an outcome scale, it's always a pain in the neck. You're never quite sure of what you're doing. I also really don't like the use of normalizing data, and then using weakly informative priors. I'm a much larger fan of, think about your prior information and restrict the feasible region to be consistent with your prior information.

Interviewer #1: Mm-hmm (affirmative). Why do you say you're not a fan, is there a specific reason?

Participant #3: I think it's too arbitrary. I think one of the real advantages of Bayesian analytics is that you can incorporate prior knowledge intelligently, and as soon as you normalize the data, and then try to assign priors to normalized data, you've lost the meaningful scale.

Interviewer #1: Mm-hmm (affirmative), I see. Okay, that's very interesting. This was great, I don't have any questions.