The modelling scenarios I usually encounter: testing complex interaction hypotheses

- **Task and dependent variable:** Participants complete a working memory or attention task that will result in button presses in response to a probe, i.e. information about whether the trials was a hit, miss, false alarm or correct rejection and a response time.
  (In my example code, the dependent variable is one that is aggregated across trials: K = (hit rate − false alarms) * set size, but what I'd really like to do is modelling single trial data to be able to fit random slopes as well. Please see a description of my problems with this below.)
- **Experimental design:** We usually have at least 2 factors, but usually 3 or 4:
  - TREATMENT (3 or more levels): e.g. different types of noninvasive brain stimulation (e.g. TMS or tACS), usually at last one nonactive condition (sham) and at least two active stimulation conditions (e.g. stimulation with a different montage, frequency or intensity etc.).
  - LATERALIZATION (2 levels): e.g. visual target stimuli appearing in the left or right visual hemifield of the computer screen, i.e. contralateral or ipsilateral to the site of neuro-stimulation.
  - Other experimental factor(s) (2 or more levels): e.g. task complexity, memory load, etc.
  - Other neuroscientific factors (often many levels): e.g. time points, frequencies, electrodes, etc.
- **Hypothesis:** We usually have interaction hypotheses, often with a specific expected pattern between Treatment and Lateralization. For example, we expect treatment A to decrease task performance (e.g. working memory capacity), but treatment B to improve task performance, relative to a control treatment. Due to the contralateral organization of the visual system, we expect any treatment effects only on trials with a target on the contralateral hemifield. So we expect:
  - Treatment A < Control Treatment < Treatment B for contralateral trials
  - Treatment A = Control Treatment =Treatment B in ipsilateral trials
  - Potentially further interactions with the additional factors

The problems I usually encounter: choosing a-priori contrasts, dependent variables & post-hoc comparisons

1. I am often a bit unsure on how to set contrasts for testing complex interaction hypotheses. For the main effect TREATMENT, it seems to me that treatment contrast coding could be quite logical. However, since this is not the only factor in the design, I usually decide against treatment coding in order to have an intercept that represents the grand average. I usually end up with sequential difference contrasts because that seems easiest to interpret for the interaction contrasts. I basically "pretend" I have an ordered factor (e.g. Treatment A < Control Treatment < Treatment B) and set sdif contrasts (e.g. Treatment A vs. Control Treatment; and Control Treatment vs. Treatment B).
   a. How feasible is this? Are there alternatives, maybe better ways to directly test interaction hypotheses?
   b. Imagine Treatment C is also included and expected to be as effective in improving task performance as Treatment B. What would be a meaningful contrast coding for such a scenario? Hypothesis:
      - Treatment A < Control Treatment **< Treatment B = Treatment C** for contralateral trials
      - Treatment A = Control Treatment =Treatment B = Treatment C in ispilateral trials
      - Potentially further interactions with the additional factors
2. My colleagues have asked for impossible contrasts (e.g. more contrasts than possible), i.e. I would have to fit multiple models or for (pairwise) post-hoc comparisons to "really know what drives the interaction".
   a. I'm not sure whether I should do any of these, considering that I can as well set contrasts a-priori and understand the interaction from these contrast estimates together with the estimated marginal means / a visualization of the estimates. Don't I have all the relevant information at hand?
   b. But if yes, how to best run such comparisons? Is it better to fit multiple models with different contrasts then or to run post-hoc comparisons?
3. Classically, most of the research that I base my studies on has used aggregated measures as dependent variables, e.g. d-prime calculated from hit rates and false alarm rates; or average reaction times calculates from single trial reaction times. What goes hand in hand with this is that by aggregating across trials, there is less data to feed into the model and I usually cannot fit random slopes, just random intercepts. So instead, I want to model single-trial data, to be able to also include random slopes. But there are a couple of issues.
   a. Single trial accuracy data: I could use a glmer model and add button press as a dependent variable and include a fixed effect RESPONSE TYPE (hit, miss, false alarm, correct rejection). Or I could add two fixed effects coding STIMULUS TYPE (target, non-target) and RESPONSE TYPE (correct, incorrect). Or I could add response times as a dependent variable and rund an lmer model. What is the best way here?

b. Single trial RTs: These single trial RTs are tremendously skewed and do not produce acceptable residuals. So instead of lmer models, I need to fit other models. Since this must be a really standard problem (response times are analysed in all kinds of disciplines, right?), I would be curious to learn about: what are practical solutions to analyzing single trial RTs? (Background: I have played with brms models with an exgaussian or lognormal family but was advised to instead run a stratified Cox model. I know very little about this and have not used these models or toolbox before (https://adibender.github.io/pammtools/articles/strata.html). And I am a bit hesitant to dive into this further because I haven't seen anyone in my field do this and I would first like to understand the question above).