Codebook:

**ID** = ID of a subject   
**D0\_S1** = 0 when the condition was David, 1 when the condition was Sarah  
**IMPR** = impression variables (Impr\_1 - Impr\_40)  
**WARM** = warmth variables (Warm\_1 - Warm\_40)  
**COMP** = competent variables (Comp\_1 - Comp\_40)  
**statval** = valence values as 1 = POSITIVE, 0 = NEUTRAL, and -1 = NEGATIVE

**statnum**

**Itemtype**

**itemnum**

**Meanval1**

**Meanval2**

|  |  |  |
| --- | --- | --- |
| |  | | --- | | **Michael Zyphur <mzyphur@gmail.com>** | | Mar 24, 2019, 2:58 PM |

|  |  |
| --- | --- |
|  |  |
|  |

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | | |  | | --- | | to Jen, Ekaterina | |   Hi Jen and Katerina,  Before the new data arrive, here are my responses to your points below:  On Wed, Feb 27, 2019 at 8:16 PM Jen Overbeck <[J.Overbeck@mbs.edu](mailto:J.Overbeck@mbs.edu)> wrote:  Hi Mike,  Katerina and I finally chatted about the desired analyses for this gender/impressions project. I’ll outline our questions below, but first a brief recap. We have so far collected about 200 responses (and plan to collect at least 1000 more). Participants went to a qualtrics survey and were told to imagine they had a new boss who was either a male or female. Then they were presented, one at a time, with 40 statements reflecting distinct things the boss had said or done. These were randomly sampled from a pool of 40 positive, 40 negative, and 20 neutral behaviors (established by pretest; and these valences break down further, so that we have positive morality, positive competence, negative morality, negative competence, etc.). After the first statement, participants were asked to report their overall impression of the new boss (1), and how competent (2) and warm (3) the new boss was. For each subsequent statement, they were shown their most recent rating and prompted to update the rating in light of the new information. At the end, they also responded to some individual difference and demographic measures.  I think of the model like this:  Y\_it = A\_t + B\_1\*Y\_it-1 + Eta\_i + E\_it  Where Y is a k=3 length vector of DVs or processes over time, A picks up intercepts, B\_1 is a matrix containing both autoregressive or AR terms and cross-lagged or CL terms capturing persistence/effects over time in the processes, Eta captures stable factors for each person, and E is 'residuals'. However, the E is better thought of as shocks or impulses, which arrive as a function of occasion-specific effects for each person:  E\_it = B\_2\*X\_it + U\_it  Where U is the actual residual for each of the three processes, X contains three additions variables to reflect a positive/negative factor x, a morality/competence factor z, and their interaction xz, and B\_2 is a matrix containing the effects of the experimental factors and their interaction.  This is a quite nice way to think about the problem, I think, because it separates the 'hangover' from the previous occasions with B\_1, while also modeling the random experimental factors at each occasion uniquely. I've never seen this in the literature, but this is the way it should be setup conceptually and Psych Methods or another journal would love to have it with a substantive example. Of course, with substitution it's easily written as a typical repeated-measures design with lagged effects included, but this is less conceptually useful:  Y\_it = A\_t + B\_1\*Y\_it-1 + Eta\_i + B\_2\*X\_it + U\_it    Our overall question is, **how do sequences of information affect impression formation differently when the target is male vs female?** We are interested in the overall impression as well as the separate warmth and competence ratings.  I'm not sure what you mean by 'target', but if you mean the participants are either M or F, then there are two ways to do this: 1) treat gender as a moderator in the model or parameter matrices directly as needed (e.g., could be via random slopes for B\_1 and B\_2 in a multilevel setup) or; 2) simple set it up as a multi-group model so each gender gets their own equation. The latter is simpler. If gender is instead an experimental factor then just treat it as a fixed or random variable (the latter if it is manipulated to randomly change over time).      For the most part, we see data analysis right now as exploratory; let me list some questions and considerations of interest:   1. Prediction: Does negative information have a more detrimental impact on overall impressions of women than of men, even when it occurs in the same or similar sequence? (We’d predict that it does.)   Unsure what this means as it seems related to the idea of a 'target'.     1. Prediction: After the introduction of a particular piece of negative information, how long does it take and what kind of subsequent information is needed to restore the impression to where it was before that introduction? We’d predict that a) for the male target, it may be sufficient to see subsequent neutral info, whereas for the female target it may be necessary to see positive info (i.e., a difference in the required *valence or intensity* of info); and b) it may be necessary to have a *greater number* of positive/neutral items to restore the pre-introduction impression for the female target than for the male.   Interesting problem. Typically B\_1 is used to model regression to the mean. B\_2 makes it a more interesting story. Together they can be used to address this question.   1. Exploratory question: Can we identify or summarize trends? Can we generate some kind of description about sequences that seem to recur? Note, the IV (info valence/order) sequences were randomly determined. We’re looking for trends in the DV…for example, do we see early fluctuations and then the impression solidifies? Or if we saw a sequence of 2+ negatives, does that produce a plunge that doesn’t recover? And finally, are the sequences different for M vs F targets?   What your saying now implies an unstable system that does not regress to any mean. That's not in the model anywhere. The best we can do is use nonlinear terms/interactions to account for what you're saying. The possibility of repeated experimental shocks to the system of various specific kinds is easy to model, but questions about non-equilibrium behaviour such as "a plunge that doesn't recover" isn't allowed in these models. Fredrickson and others don't seem to understand this about typical statistical models (here's a recent example of this misunderstanding by [Banaji](https://journals.sagepub.com/doi/abs/10.1177/0956797618813087)). Alternative methods are needed. Long-run trends can be accounted for using econometric methods for integrated time series, or the EDM method I talked anout at SOBA can be used. I doubt that your data will exhibit the kinds of patterns required. I'll bet your series are mean reverting and stable/non-complex.   1. Note: If 40 time points is too much, we could define smaller sequences. We’ll follow your guidance on this. One thing that could be nice: We could randomly sample shorter sequences from the dataset, do exploratory analyses, and generate predictions to test with other randomly-sampled sequences from the same dataset. (Then of course we expect to test them in a confirmatory manner with the new N=1000 dataset.)     More is always better.   1. Note: I’m attaching three versions of the dataset. The first is absolutely raw data from Qualtrics. The second was restructured by Katerina to show which statement was shown, for each participant, in positions 1 through 40. The third was created by an RA of mine to show the ‘overall impression’ rating made by each participant in positions 1 through 40. Thus, matching the 2nd and 3rd dataset would give you, organized by order 1 through 40, the statement made and the overall impression that resulted. [Edit: I just visually checked the files and found that my RA didn’t organize the ratings in the same 1-40 order, but rather by statement. I’ve asked her to redo the file but am sending you this version meanwhile, in case it’s usable.] Hopefully this data structure (or at least the raw dataset) will give you something useful. I also attach the Qualtrics file in case you need it for reference, though you can also point out any deficiencies and I can have my RA fix them.   I'll wait for the new data.     1. Question for you: It could be informative to weight the statements by the intensity of valence. That is, not only knowing that a statement was negative (“told a racist joke”) but that it was *really*negative (e.g., compared with the milder “didn’t clean out the refrigerator”) could tell us more about how the impressions develop. We don’t currently have the ratings needed to do such weighting, but we can run a separate study to collect them, and we could do that while you poke around with the attached datasets. Will it be possible to add weighting or scaling? Put differently, should we go ahead and collect ratings data?   If you had such data there would be various ways to incorporate them into the analyses. It's up to you.    Thanks so much for joining us—we’re really excited to see what you can find. And obviously, don’t hesitate to ask questions of either or both of us.  Too easy. Sorry for the wait. Life exploded temporarily but is now returning to normal.      Cheers,  Jen |