Supplemental Material
for
A mental models study of hurricane forecast and warning production, communication, and decision making

Supplement A
Group Session Protocol and Resulting Diagram
Thanks for agreeing to help us out. Our main goal today is to produce an expert model of the hurricane forecast and warning cycle that describes the decisions in the cycle. And as you know, we’re focusing on hurricanes and hurricane warnings in the Miami-Dade area.

Part I. [45 minutes]
We’re going to start by asking you to list all the possible pieces you can think of for the hurricane forecast and warning cycle. To make sure we’ve covered the entire territory, we have a number of specific questions and prompts we’ll go through that will sound familiar to all of you. Please think out loud and discuss among yourselves, writing each piece you think of on a sticky note as you go, one piece per sticky note, and we’ll put them up on the wall. Write down everything you think of, no matter when it occurs to you. We can always add something later.

1. What do you think determines whether a hurricane impacts Miami-Dade or not?
2. What risks are there from hurricanes in Miami-Dade? What are the possible consequences of hurricanes in Miami-Dade? Who is at risk from hurricanes in Miami-Dade? What are the possible consequences of hurricane warnings in Miami-Dade?
3. What can or should be done, if anything, to reduce risks from hurricanes in Miami-Dade?
4. What information do people need to protect themselves from hurricane risks in Miami-Dade? How can a person find out if there is a general risk of a hurricane at a specific location in Miami-Dade? How can a person find out if there is a risk of an approaching hurricane where they are?
5. What are your goals for hurricane warnings? What are you trying to achieve when you issue a warning?
6. What are your decision rules or criteria for issuing a hurricane warning? How do you decide whether or not to warn, when to warn, and for what area to warn?
7. What is the most important information or outside input you use in deciding to warn? How is the warning put together? What content do you include in a hurricane warning? Anything about uncertainties?
8. Once you have created a warning, how and to whom is it disseminated? What, if any, are the roles of intermediaries (e.g., media, public officials) in the dissemination of warnings, and how do you interact with them? Who uses hurricane warnings in Miami-Dade? How do they use them?

9. What feedback do you get about the hurricane warnings you issue? How could warnings be improved to better meet users’ needs?

**Part II. [27 minutes]**

Now we would like to obtain a complete picture from you of the forecast and warning cycle for hurricanes in Miami-Dade, from generation to use of the forecast and warning, and consequences, including any important feedback loops. To get you started here are the diagrams you each constructed individually.

We’d like you to use the key factors that you’ve discussed so far and in light of the ideas from the individual diagrams create a consensus influence diagram of forecast and warning decisions, including the outcomes and consequences of hurricane warning decisions. Please talk through your rationale for the placement of the pieces and be sure to address areas where there is disagreement.

[Prompt:
- Is this all of the key factors that influence forecasters’ hurricane warning decisions?
- Please expand your diagram to address people’s decisions using hurricane warnings, including evacuation decisions and their consequences, by revisiting the list you have of what factors might influence people’s decisions to take action or not if there is a hurricane warning.
- Imagine a Hurricane Floyd size hurricane is predicted to make landfall in 5 days. Walk through the diagram and tell us how it describes what will happen. Revise the diagram as you see fit.
- Now imagine that a Hurricane Andrew size hurricane is predicted to make landfall in [5 days, 3 days, 2 days, hours] Walk through the diagram and see how these times affect what will happen. Revise the diagram as you see fit.]

What is the relative importance of each part of this model for determining the quality of the forecast and the risk reduction potential for the forecast-warning cycle? In other words, what contributes to a better forecast or warning?

Where is there room for improvement in the forecast and warning cycle now? Where are the greatest uncertainties?

**Part III. [30 minutes]**

Another of our research goals is to get a better sense of some of the parameters involved in the forecast cycle. We’d like you to estimate several probabilities that we suspect are key. For these, please consider the full distribution of hurricanes that could occur, of all sizes and intensities. Wed also like you for each probability to give us your best (mean) estimate, a lower bound, and an upper bound.

1. How likely do you think a hurricane is to occur in Miami-Dade in the next year?
2. If a hurricane warning is issued for Miami-Dade, how likely is it that hurricane conditions will occur in Miami-Dade during the warning period?

3. If a hurricane watch is issued for Miami-Dade, how likely is it that hurricane conditions will occur in Miami-Dade during the watch period?

4. If a hurricane intensifies into a stronger storm, how likely is it that this intensification was forecast?

5. If a hurricane weakens into a weaker storm, how likely is it that this weakening was forecast?

6. If a hurricane shifts directions away from Miami-Dade, how likely is it that this shift was forecast?

7. If a hurricane increases in translational speed, how likely is it that this increase was forecast?

8. If there are tropical storm-force winds in Miami-Dade now, how likely is it that a hurricane warning was issued 36 hours ago?

9. If a hurricane warning is issued for Miami-Dade, how likely is it that any given person in the warned area will receive the warning?

10. If a hurricane warning is issued for Miami-Dade, how likely is it that any given person in the warned area will evacuate?

Part IV. [10 minutes]
We are also interested in your perspective on different components of information about hurricane risks and potential improvements in the communication and use of this information.

1. Please describe how information in hurricane watches and warnings convey the risks associated with winds, storm surge, tornadoes, inland flooding, and other aspects of hurricane risks.

2. How do you think this information could or should be presented in new or different manners to provide more useful information for decision making?

Part V. [5 minutes]
Recognizing that everyone has their own specialized expertise, it would nevertheless be helpful for us to hear your judgments about the expertise in your office related to our research project.

1. First, it would be helpful to know what single person in your office you would select to best estimate the probabilities we just discussed for the forecast and warning cycle for hurricanes. Who would you select?

2. Would the same person in your office best describe the structure of the hurricane forecast and warning cycle overall? If not the same person, who? Would you be willing to let us send you the model for additional comments you might have sometime in the near future?
Group Diagram (dotted lines indicate concepts for which the group participants did not indicate a shape)
Supplement B

Coding Methodology and Inter-Coder Reliability

Development of Coding Scheme

We developed the initial coding scheme based on the diagram forecasters in the group session created. All the factors identified in the group model were listed in a spreadsheet and arranged according to the hierarchy apparent in the group diagram. We then coded each of the influence diagrams developed in the forecasters’ individual mental models interviews into the coding scheme. When individual diagrams contained topics that were subsets of a code in the existing coding scheme or were not in the group diagram, these were added to the decision model and coding scheme.

Next, we developed a written protocol for applying codes from this scheme to the interview transcripts, according to which we coded each individual interview transcript. New concepts mentioned by an interviewee could be accommodated by adding or amending codes. The resulting scheme (after coding all of the individual professional interviews) comprises the decision model depicted in Figures 1 and 2 of the manuscript.

Coding Process

Prior to beginning the formal coding, three coders coded the first NHC interview, to train in the coding methods. Next, two coders independently coded the first interview of each group of professionals (NHC forecasters, WFO forecasters, public officials, and broadcasters). We then calculated inter-coder reliability statistics to identify discrepancies, following which the coders discussed each question in each of the coded interviews, to examine agreement and disagreement on codes and to revise the coding scheme as needed to accommodate novel responses. After this initial round(s) of coding, the two coders re-coded the same interviews independently, using the revised coding scheme. These same two coders then also coded all other NHC and WFO interviews. For the public official and broadcaster interviews, only the first interview in each group was coded by both coders; one of the two coders coded the remaining public official and broadcaster interviews.

Inter-coder Reliability Calculation

Reliability statistics were calculated for all of the NHC and WFO forecaster interviews and for the first public official and broadcaster interviews. For the inter-coder reliability calculations, the coding results were output in a binary format, with each item in the coding scheme coded “1” if it occurred and “0” if it did not occur in an interview transcript. Because responses to each question could be assigned any code in the coding scheme, each response (and thus each interview section) could be coded “1” or “0” for up to the full 189 codes. For example, if both coders coded a response as code 11100 Intensity (wind speed), it would be coded as an agreement (a match) on that code. Absence of code 11130 Winds greater than 64 KTS/74 MPH - hurricane designation in both codings would also be coded as agreement (both coded as “0”).
Inter-coder reliability (see Table B1) was calculated using Deen G. Freelon’s inter-coder reliability web service, ReCal (Freelon, 2010)¹ using two output statistics: a) average pairwise percent agreement between two coders and b) Cohen’s Kappa, which is an agreement statistic that corrects for the likelihood of agreeing by chance. Percent agreement between two coders for each interview was calculated by counting codes on which both coders agreed (both had 1 or both had 0) and dividing by the total number of codes in the coding scheme. Cohen’s Kappa was calculated by cross-tabulating frequency counts for agreement on each code category (e.g., both coding “1” for a given code category) and calculating observed and expected agreement (see e.g., Neuendorf 2002).

For each of the interviews (with the exception of the one incomplete NHC interview), we calculated inter-coder reliability for the full interviews. Because the second NHC interviewee was only asked questions in the “hurricane experience” section, reliability was calculated only for this question type on that interview.

For the first interview in each group, inter-coder reliability was calculated for the coding before the inter-coder discussion as well as after. As expected, inter-coder reliability for each interview increased when the transcripts were recoded after the discussion: for the first WFO forecaster, from 79.3% agreement to 85.7% (Cohen’s Kappa from 0.56 to 0.71); for the first public official, from 80.3% to 86.8% (Cohen’s Kappa from 0.59 to 0.73); and for the first broadcaster, from 85.6% to 92.1% (Cohen’s Kappa from 0.67 to 0.82). In sum, for the full interviews, reliability varied from 75.3% to 92.1% (percent agreement) and from 0.49 to 0.82 (Cohen’s Kappa from 0.48 to 0.82).

| TABLE B1. Percent agreement and Cohen’s Kappa for full interviews. |
|-----------------|------------------|
|                 | Percent agreement | Cohen’s Kappa |
| NHC1            | 80.95%            | 0.620         |
| NHC2            | 78.13%            | 0.479         |
| NHC3            | 75.31%            | 0.494         |
| NHC4            | 78.84%            | 0.578         |
| WFO1            | 85.71%            | 0.707         |
| WFO2            | 80.95%            | 0.609         |
| WFO3            | 87.30%            | 0.736         |
| WFO4            | 79.37%            | 0.592         |
| PO1             | 86.77%            | 0.729         |
| BR1             | 92.06%            | 0.821         |

References