Dr. Levine discussed concerns with experimental design and data interpretation. First, she talked about different ways of gathering data. Observational studies simply collect data that is already available. Dr. Levine gave the example of a study that observed Olympic medal ceremonies and found that bronze medalists smiled more than gold and silver medalists. However, observational studies have no control over the circumstances under which the data is collected, so it can be difficult to draw conclusions from it. In the Olympic medalists study, there is no way to know if the bronze medalists actually felt happier than the gold and silver medalists - perhaps they were disappointed but putting on a brave face. This particular example could be dealt with by applying Paul Ekman's work on facial actions. Ekman and his collaborators isolated the muscles of the face and defined combinations of muscle movements that corresponded to facial expressions. Among other things, Ekman's system is able to distinguish between smiles that reflect genuine happiness and those that do not.

Another method of gathering data is conducting surveys. There are several factors that affect people's responses to surveys: the wording and order of the questions, the scaling of the possible responses, and the way in which the survey is administered (in person or online). For example, asking "How often do you…" causes people to overestimate how often they do things; asking "How infrequently do you…" causes people to underestimate. Dr. Levine mentioned that one way to avoid inadvertently biasing a survey is to randomize the wording and order of the questions. However, this is not always possible due to relationships among the questions – some questions need to be asked in order so that they make sense.

It seems to me that we could randomize the order of chunks of questions while keeping together the more closely related questions. We could even create a tool to automatically randomize surveys. At least with the "how often"-"how infrequently" example, it would be straightforward to automatically identify and replace such expressions with their opposites. For reordering questions, it may be possible to apply topic segmentation or discourse relations – or a combination of both – to find chunks of questions that need to stay in a particular relative order. I am not convinced that either would be enough by itself. Topic segmentation may be too fine-grained; because questions on a survey are all asking fairly different things (to do otherwise would create redundancy), topic segmentation may assign to each question a different topic, especially since it would need to make its decision based on the words present in each question. Discourse relations could be difficult to identify from the questions because there would not be any explicit discourse connectives (adverbs and conjunctions). Also, detecting relations among questions could require outside knowledge about the purpose of the survey.

Throughout the class, Dr. Levine illustrated her points with examples given by students. While discussing one such example, she mentioned that how someone behaves after being told to lie and how he behaves after being told to say a predetermined, untrue statement would be different, and so the two should be separate experiments. This makes sense, but to establish the baseline for a polygraph, the subject is told to tell a specific lie, even though the goal of the polygraph is to catch the subject telling a spontaneous lie. I wonder if any research has been done on the differences between the two types of lies. This would seem to be a very important point to clarify, given the potential consequences of faulty polygraphs.

The last topic that Dr. Levine discussed was that of ethics. I asked if it would be unethical to mislead people about the study in which they were participating. My undergraduate senior project was on improving review quality in online marketplaces. My partner and I conducted an experiment on the Amazon Mechanical Turk in which participants played the ultimatum game. They were told that they were randomly matched with another participant to play ten games, after which they were asked to rate their partners. In fact, the partners were bots, and the focus of the experiment was not the ultimatum game but the ratings that the participants left. It could be easily argued that my partner and I lied to the participants. Dr. Levine responded that, in this case, she thought that what we did was alright. There was no way that we could have told the participants what our experiment was really about without affecting our results. As long as the participants were aware of what they would be asked to do, were debriefed after they finished the experiment, and came to no harm as the result of our misleading them, it would be fine not to tell them the whole truth.