Seminar in Cognitive Science
DQ #5

Ozaner Hansha

February 25, 2021

Question 1

Prompt: What is the belief inhibition hypothesis? Explain the type of task Leslie and Polizzi (1998) used to test it and what they showed.

Response: The belief inhibition hypothesis posits that, for children to be able to have accurate theories of mind, they must be able to inhibit their own beliefs so that they might be able to reason about the actions of others who have different beliefs (e.g. why Sally checks the basket in the false-belief task).

Leslie and Polizzi tested a group of 4-year-olds that passed the standard false-belief task. Half of them were given a true belief task (e.g. Sally knows that the ball was moved) but the task had them shift desires (e.g. Sally wants to check where the ball *isn't*). The other half were given a false belief task (e.g. Sally doesn't know that the ball was moved) and again had them shift desires.

As expected, the first half almost entirely passed the test, showing that adding the shifting of desires to the experiment shouldn't be so complicated to a 4-year-old that it skews the results. The second half, however, only passed the task with a 38% accuracy. Bear in mind that these children had already passed the non desire shifted false-belief task before.

Leslie and Polizzi explained this by hypothesizing an *inhibition of return*. Essentially, the child's beliefs are inhibited to account for the false-belief task, but then they have to be negated to deal with the switched desire. But this negation would return the child's answer to their original non-inhibited answer which they are averse too.

Question 2

Prompt: What is the basic learning problem that research on theory of mind addresses?

Response: The fundamental learning problem that a child faces in developing a theory of mind is the following: if a child cannot see, hear, feel, or otherwise sense mental states, how are they to learn what they are/how they work and how they can be used to predict other's behavior? Indeed, adults with developed theories of mind posit these immaterial *minds* and *beliefs* to certain objects like people and can reason about their behavior in this way. For a child to fully be able to predict another's behavior, they must be able to learn of these beliefs and their properties without ever directly observing them.

Question 3

Prompt: What are the two main components of the Creative Adversarial Network (CAN)? Explain (loosely) how these components interact to create art.

Response: The CAN architecture is comprised of two main components: the discriminator and the generator. The discriminator is given and trained on a large dataset of art all labled with whatever style they fall under, ultimately being able to discriminate between these different art styles. The generator, blind to the art dataset given to the discriminator, is trained to generate art pieces by receiving from the discriminator. That input is comprised of two signals: the first is a measure of how well what was generated can be called 'art' rather than 'not art', leaving out what particular style the discriminator thinks this art fall under. The second input is a measure of how well what was generated falls into a specific style of art the discriminator has seen.

By maximizing the first signal (i.e. being art), and minimizing the second (i.e. making images that don't fit into a certain style), the generator is forced to generate art that is necessarily novel, or creative.

Question 4

Prompt: What insights might be gained from computational approaches to artistic style? What are the challenges of this approach?

Response: Computational approaches to art might give rise to new styles not yet explored by humans, and may also allow for a deeper understanding of human perception of art and its qualities in a slightly more objective way.

Some challenges these approaches face are: it is hard to measure success or failure in the art world and that methods used in deep learning are notoriously obfuscated which makes it hard to gather insight about what the computer is doing.

Question 5

Prompt: What were the major findings in the analysis of art style? What patterns were able to be extracted from the paintings?

Response: One interesting finding was that all the models' learned representations showed a smooth transition over time, despite the models not being given any information on dates/authors of the works. This is exactly what one would expect given that art styles throughout history follow from each other and evolved over time.

Another interesting finding is that the learned representations seem to correlate with 'Wolfflin's concepts' a set of categorizations used to classify art by its features. In particular the categories of 'planer vs recession' and 'linear vs. painterly' seemed to fit quite well the PCA of the models' categorizations. This gives credence to Wolfflin's criteria in an empirical way. However we should temper this by noting that not all of Wolfflin's concepts were well correlated with the linear models.

Finally a third finding was that, not only were the models clearly able to identify different art styles, but analyzing their output leads for a very reasonable notion of 'representative artists' of particular art styles in terms of the variance of their work with other works in the category.