

Preregistered Hypotheses for Evaluating Models of Literary Character

November 5, 2013

1 Hypotheses about individual characters.

We're preregistering specific hypotheses about the similarities of individual characters in order to test our model. 30 hypotheses total: 5 strong, 10 moderate, and 15 weak hypotheses.

1.1 Strong hypotheses.

Sanity checks to start with. Pairs of very similar characters ought to resemble each other more than they resemble characters drastically different in every way. If we get any of these three wrong, something must be wrong with our math.

1. Elizabeth Bennet in *Pride and Prejudice* resembles Elinor Dashwood in *Sense and Sensibility* (Jane Austen) more than either character resembles Allen Quatermain in *Allen Quatermain* (H. Rider Haggard).
2. Sherlock Holmes, in *The Adventures of Sherlock Holmes*, resembles Holmes in *The Sign of Four* (Conan Doyle) more than he resembles Bill Sikes in *Oliver Twist* (Dickens).
3. Emily St. Aubert in *The Mysteries of Udolpho* resembles Adeline in *The Romance of the Forest* (both by Ann Radcliffe) more than either character resembles Uriah Heep in *David Copperfield* (Dickens).

Slightly more challenging. Characters who reappear in multiple works ought to resemble themselves more than they resemble other characters in the same books.

4. ~~Natty Bumppo in *The Pathfinder* (James Fenimore Cooper) resembles Bumppo/the scout/Hawkeye in *Last of the Mohicans* more than either version of the character resembles Uncas in *Last of the Mohicans*.~~ The character names we expected were not identified by the NLP pipeline, so we didn't use this for model evaluation. (The problem is possibly that Bumppo is called "The Pathfinder," which may not have been recognized as a proper name.)
5. Sherlock Holmes, in *The Adventures of Sherlock Holmes*, resembles Holmes in *The Sign of Four* (Conan Doyle) more than either version of Holmes resembles Watson in *Adventures*.

1.2 Moderate hypotheses.

Character types ought to be recognizable across the works of a single author. We start with cases that should be pretty clear: similar characters in the same author should resemble each other more than they resemble slightly different characters in a different author. These should be easy cases in a naive model. But these apparently easy cases might become harder if we attempt to “factor out” authorship or date of publication. In other words, these hypotheses are here to guard against overcomplicating the model.

6. Emily St. Aubert in *The Mysteries of Udolpho* resembles Adeline in *The Romance of the Forest* (both by Ann Radcliffe) more than either character resembles Elizabeth Bennet in *Pride and Prejudice* (Austen).
7. Wickham in *Pride and Prejudice* resembles Willoughby in *Sense and Sensibility* (Jane Austen) more than either character resembles Mr Rochester in *Jane Eyre* (Brontë).
8. Lucy Snowe in *Villette* resembles Jane Eyre in *Jane Eyre* (Charlotte Brontë) more than either character resembles Pip in *Great Expectations* (Dickens). (All three first-person narrators.)
9. Diana “Die” Vernon in *Rob Roy* (Walter Scott) resembles Flora MacIvor in *Waverley* (Scott) more than she resembles Emily St. Aubert in *The Mysteries of Udolpho* (Radcliffe). (Diana and Flora are active characters surrounded by mystery; Emily St. Aubert is a largely passive point-of-view character.)

Now we make this a bit harder by contrasting pairs of similar characters to different characters in the same author.

10. Wickham in *Pride and Prejudice* (Jane Austen) resembles Willoughby in *Sense and Sensibility* more than either character resembles Mr Darcy in *Pride and Prejudice*. (Can we identify unreliable seducers?)
11. Elizabeth Bennet in *Pride and Prejudice* (Jane Austen) resembles Elinor Dashwood in *Sense and Sensibility* more than either character resembles Mrs Bennet in *Pride and Prejudice*. (Can we recognize thoughtful protagonists?)
12. John Barton, in *Mary Barton* (Elizabeth Gaskell) resembles Nicholas Higgins in *North and South* more than either character resembles Margaret Hale, in *North and South*. (Can we recognize honest workingmen?)
13. Agnes Wickfield in *David Copperfield* (Charles Dickens) resembles Lizzie Hexam in *Our Mutual Friend* more than either character resembles Uriah Heep in *Copperfield*. (Self-denying young women with troubled romantic histories.)
14. Rowena in *Ivanhoe* (Walter Scott) resembles Rose Bradwardine in *Waverley* more than she resembles Rebecca in *Ivanhoe*. (A slightly debatable hypothesis, but this at least is Alexander Welsh’s argument about Scott’s “dark” and “light” heroines.)

15. Philippe, Marquis de Montalt in Ann Radcliffe's *Romance of the Forest* resembles Montoni in *The Mysteries of Udolpho* (Radcliffe) more than either character resembles Valancourt in *Udolpho*. (Gothic villains resemble each other more than they resemble the love interest.)

1.3 Weak hypotheses.

Comparing characters across points of view is a “weak hypothesis” in part because it's not clear that we can fully separate character from the point of view used to depict it. I've selected cases that should be relatively clear even if we assume that these variables are entangled: strongly similar characters in the same author should resemble each other more than they resemble characters who are quite different.

16. Pip, in *Great Expectations* (Dickens) should resemble Oliver in *Oliver Twist* more than either character resembles Miss Havisham in *Great Expectations*. (Both youthful protagonists in coming-of-age stories, but Pip is also a first-person narrator.)
17. Pip, in *Great Expectations* (Dickens) should resemble Oliver in *Oliver Twist* more than he resembles Allen Quatermain in *Allan Quatermain* (H. Rider Haggard). (Here the test is that *Quatermain* is also a first-person story, but narrated by a character very different than Pip—a grizzled explorer.)
18. Frank Osbaldistone in *Rob Roy* resembles Edward Waverly in *Waverly* (Walter Scott) more than either character resembles Rashleigh Osbaldistone in *Rob Roy*. (Two Scott protagonists should resemble each other more than they resemble a scheming villain, even though one of the protagonists is a first-person narrator.)
19. Huckleberry Finn, in *The Adventures of Huckleberry Finn* (Mark Twain), resembles Huck Finn in *Tom Sawyer* more than either version of Huck resembles Tom in the latter book. (This is actually nontrivial, although Huck is “the same person,” because they're very different books with different kinds of narration.)

Now we begin to make comparisons between similar characters in *different* authors. Here it is frankly difficult to find uncontroversial cases. We start with comparisons between three different authors, which are slightly less challenging than cases where we're asking the model to separate characters written by the same author.

20. Ralph Rover, narrator of *The Coral Island* (R. M. Ballantyne), resembles William, the narrator of *Swiss Family Robinson* (J. D. Wyss, in translation) more than either character resembles Long John Silver in *Treasure Island* (R. L. Stevenson). (A fairly straightforward case, since the first two books are Robinsonades, and the third isn't, in spite of its island setting.)
21. Alice Humphreys, in *The Wide, Wide World* (Susan Warner) resembles Helen Burns in *Jane Eyre* (Charlotte Bronte) more than either character resembles Simon Legree in *Uncle Tom's Cabin* (Harriet Beecher Stowe). (Spiritually-minded, sickly guide/helper figures resemble each

other more than they resemble a brutal slaveowner. Should not be hard unless some aspect of UK/US diction throws things off.)

22. Bill Sikes in *Oliver Twist* (Dickens) resembles Long John Silver in *Treasure Island* (Stevenson) more than either character resembles Francisco in *The Pirate* (Frederick Marryat). (Sikes and Silver are both villains, and Francisco is more or less the hero. But this is a relatively challenging test, because that characterological similarity is confounded by similarities of period (Dickens and Marryat are contemporary) and similarities of nautical setting (*The Pirate / Treasure Island*) that could shorten the other two legs of the triangle.)
23. Montoni in *Mysteries of Udolpho* (Radcliffe) resembles Heathcliff in *Wuthering Heights* (Emily Brontë) more than either character resembles Mr Bennet in *Pride and Prejudice*. (Another case where we're testing our model's ability to discern similarities across time. Montoni and Heathcliff are both cruel, proud, tormented figures who get called "Gothic" or "Byronic," although Heathcliff is more sympathetic. Mr Bennet is nothing like them morally, but he is intermediate between them in time; if the model cares too deeply about period diction, it might be thrown off.)
24. Arthur Donnithorne, in *Adam Bede* (George Eliot) resembles Alec d'Urberville in *Tess of the d'Urbervilles* (Thomas Hardy) more than either character resembles John Thornton in *North and South* (Elizabeth Gaskell). (The first two characters are aristocrats who seduce a young working-class woman; the third character is a mill-owner and a worthy suitor.)

Now we have comparisons where we're asking the model to understand that two characters in different authors can be more similar than two in the same author. This is a challenging test, and also perhaps based on debatable critical premises—since it's not clear how far human readers are able to separate modes of representation from the thing represented. But we have explicitly flagged these as weak hypotheses; take them with a grain of salt.

25. Jonathan Wild in *Jack Sheppard* (W. H. Ainsworth) resembles Bill Sikes in *Oliver Twist* (Dickens) more than either one resembles Mr. Brownlow in *Oliver Twist*. (Brutal murderers in the Newgate genre resemble each other more than they resemble a philanthropist; should be straightforward unless the model is allowing authorship to trump character.)
26. Simon Legree in *Uncle Tom's Cabin* (Harriet Beecher Stowe) resembles Bill Sikes in *Oliver Twist* (Dickens) more than either of the characters resemble Tom in *Uncle Tom's Cabin*. (Legree and Sikes are both violent and coarse; the test here is whether that similarity can overcome the distance of time and authorship.)
27. Isabel Archer, in *Portrait of a Lady* (Henry James) resembles Gwendolyn Harleth in *Daniel Deronda* (George Eliot) more than either character resembles Madame Merle in *Portrait*. (Both Isabel and Gwendolyn are independent spirits trapped in marriages to manipulative husbands. James is said to have drawn explicitly on Eliot's novels (and perhaps specifically on Harleth) as models for Archer. Edel, Leon. *Henry James: The Conquest of London: 1870-1881*. Philadelphia and New York: J. B. Lippincott Company, 1962. However, no comparison involving James is going to be easy. I would expect the model to start failing here.)

28. Gilbert Osmond, in *Portrait of a Lady* (Henry James) resembles Henleigh Mallinger Grandcourt in *Daniel Deronda* (George Eliot) more than either character resembles Daniel Deronda in *Deronda*. (Our working hypothesis is that cruel, manipulative husbands resemble each other more than they resemble stand-up guys.)
29. Silas Ruthyn, in *Uncle Silas* (Sheridan Le Fanu) resembles Montoni in *The Mysteries of Udolpho* (Ann Radcliffe) more than either character resembles Valancourt in *Udolpho*. (Gothic villains who play a similar role in the plot, but are located at the rather challenging distance of a century from each other. This will be very hard.)
30. David Copperfield in *David Copperfield* resembles Jane Eyre in *Jane Eyre* more than either character resembles Mr Rochester in *Jane Eyre*. (The slightly debatable premise here is that the protagonists of a Bildungsroman resemble each other, in spite of gender differences, more than they resemble a Byronic love interest.)