WHY DO PEOPLE WRITE FANFICTION?

Using computational methods to study reader preferences

Mia Jacobsen
Research Assistant, Center for Humanities Computing
mia@cas.au.dk

CENTER FOR
HUMANITIES
COMPUTING

Admiration and Frustration: A Multidimensional Analysis of Fanfiction

Mia Jacobsen[1,*], Ross Deans Kristensen-McLachlan[1,2]

[1]*Center for Humanities Computing, Aarhus University, Denmark*
[2]*Department of Linguistics, Cognitive Science, and Semiotics, Aarhus University, Denmark*

Abstract

Why do people write fanfiction? How, if at all, does fanfiction differ from the source material on which it is based? In this paper, we use quantitative text analysis to address these questions by investigating linguistic differences and similarities between fan-produced texts and their original sources. We analyze fanfiction based on *Lord of the Rings*, *Harry Potter*, and *Percy Jackson and the Olympians*. Working with a corpus of around 250,000 texts containing both fanfiction and sources, we draw on Biber's Multidimensional Analysis [4], scoring each text along six dimensions of functional variation. Our results identify both global and community-based preferences in the form and function of fanfiction. Crucially, fan-produced texts are found not to diverge from their source material in statistically meaningful ways, suggesting that fans mimic the writing style of the original author. Nevertheless, fans as a whole prefer stories with less focus on narrative and greater emphasis on character interactions than the source text. Our analysis supports the notion proposed by qualitative studies that fanfiction is motivated both by admiration for and frustration with the canon.

A SHORT NOTE ON VOCABULARY

Fandom = A group of fans centered around a specific fan object
- Fan objects can be anything from TV-shows and books to real life people

Fanfiction = Piece of fan-written fiction that centers around the characters or people from the fan object
- Also abbreviated as 'fanfic' and 'fic' → I will use these interchangably

MOTIVATION AND BACKGROUND

WHY DO PEOPLE WRITE FANFICTION?

- Research on fans and fanfiction is a traditionally qualitative and ethnographic endeavor

- In recent years, computational analyses of fanfiction have gained traction

- However, these studies are focused on the texts as texts and less so on the people who produce it
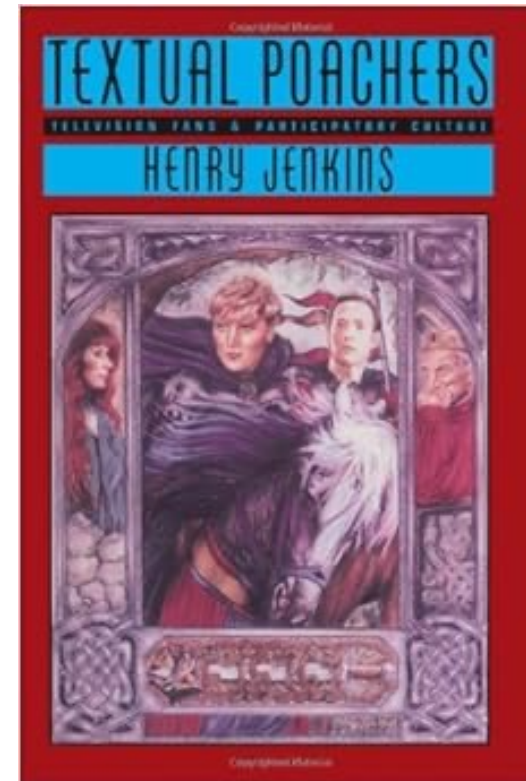
SCHOOL OF CULTURE AND SOCIETY

AARHUS UNIVERSITY

WORKSHOP
11 JUNE 2025

MIA JACOBSEN
RESEARCH ASSISTANT

5

" fan fiction loses its meaning if removed from its context. Fan fiction thus offers insight into the fan community – its conversations, its tropes, and its members' discussions and concerns

\- BUSSE, 2017

Research question: What can a large-scale analysis of fanficiton tell us about the motivation behind fanfiction writing?

CORPUS COLLECTION AND FEATURE EXTRACTION

WHICH TEXTS?

A few starting criteria

- Text-based fandoms
- Well-established fandoms
- Source texts from the same genre
- No *crossovers* and only texts in English

Archive of Our Own

1 - 20 of 485,365 Works in Harry Potter - J. K. Rowling

← Previous 1 2 3 4 5 6 7 8 9 … 4999 5000 Next →

1 - 20 of 37,125 Works in Percy Jackson and the Olympians - Rick Riordan

← Previous 1 2 3 4 5 6 7 8 9 … 1856 1857 Next →

1 - 20 of 24,898 Works in The Lord of the Rings - J. R. R. Tolkien

← Previous 1 2 3 4 5 6 7 8 9 … 1244 1245 Next →

WHICH METHOD OF ANALYSIS?

Biber's Multidimensional Analysis

- 6 dimensions of *functional variation* in the English language

- These dimensions are based on groups of grammatical-semantic linguistic features

- A well-established method for comparing registers and genres

- Creates an explicit link between form and function
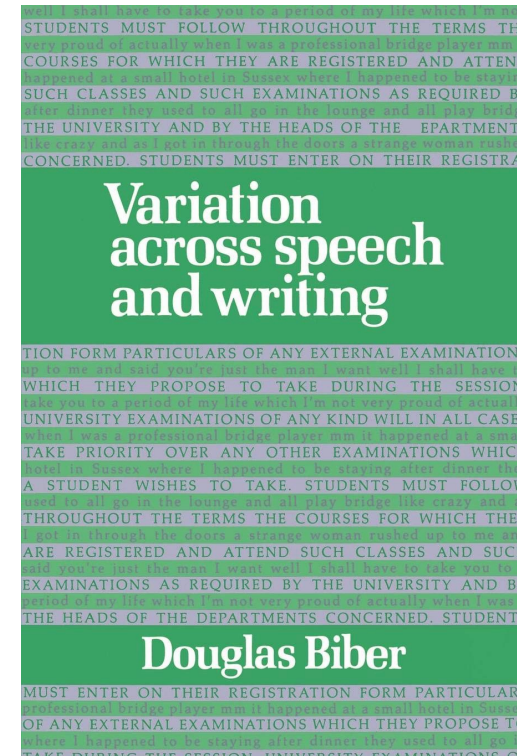
Variation across speech and writing

Douglas Biber

ANDREA NINI'S MULTIDIMENSIONAL ANALYSIS TAGGER

Table 2
Summary of dimensions of variation established using MDA

|	Summary	Short Description
Dimension 1	Involved versus Informational	*Informational*: Dense and careful information integration. *Involved*: Verbal style with a focus on the here-and-now.
Dimension 2	Narrative Concern	Distinguishes between texts with a narrative focus from others
Dimension 3	Context-(in)dependent referents	*Context-dependent*: Receiver must use context to infer what time and place is being referred to. *Context-independent*: The referents in the text are made explicit and thus not dependent on the context
Dimension 4	Overt expression of persuasion	The degree to which the sender's opinion is overtly expressed and/or overt attempts to persuade the receiver are made
Dimension 5	Abstract style	Distinguishes between informational discourse that is abstract and technical from informational discourse that is not
Dimension 6	On-line information elaboration	Indicates whether the information presentation is made carefully or is more fragmented

500-token snippet

↓

Tagged using MAT

↓

Dimension scores calculated for each text
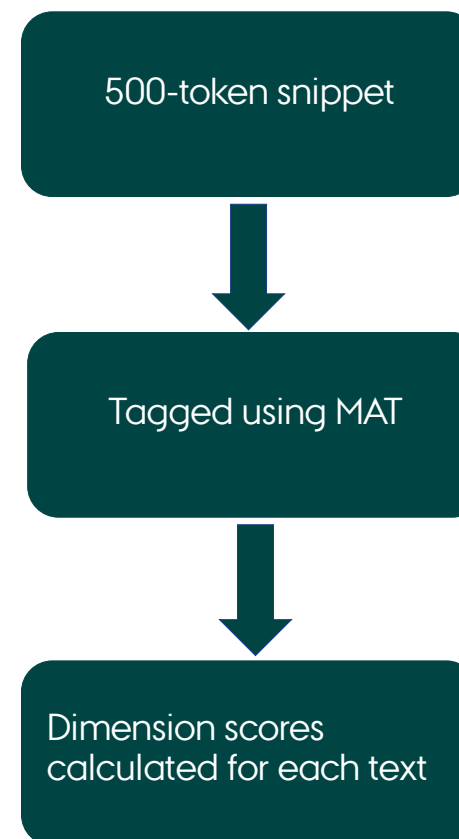
THE FINAL CORPUS

Table 1

Corpus summary

	Harry Potter	Percy Jackson	LOTR	Total
Fanfiction	224,039	17,823	12,523	254,385
Source texts	202	69	84	355
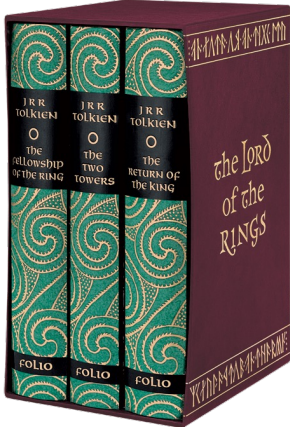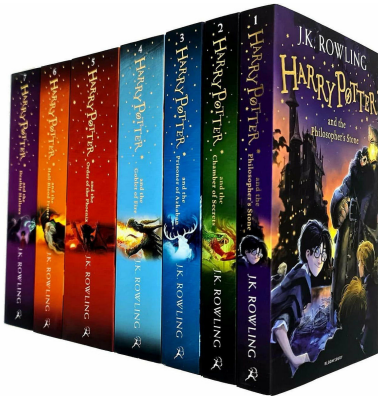
STATISTICAL ANALYSIS

Three questions:
1. How, if at all, do fanfics differ from their source material?
2. How are fans' reading preferences related to the dimension scores across groups?
3. What does that tell us about the motivation for writing fanfiction?

Answered using linear mixed effects models
- Explicitly models the fact authors might be repeating
- Can account for imbalance in the data
- Allows for modeling of *interaction effects*

STATISTICAL ANALYSIS

How, if at all, do fanfics differ from their source material?

$$Dimension \sim text\ type + fandom + (1|author)$$

Weights were added because of the imbalance in the data

STATISTICAL ANALYSIS

How are fans' reading preferences related to the dimension scores across groups?

$$Dimension \sim engagement * fandom + published\ date + word\ count + (1|author)$$

=> Engagement metric = $\frac{kudos}{hits} * 100$

RESULTS

RESULTS

No statistical differences across text types

However, fanfics look more similar than source texts

SCHOOL OF CULTURE AND SOCIETY
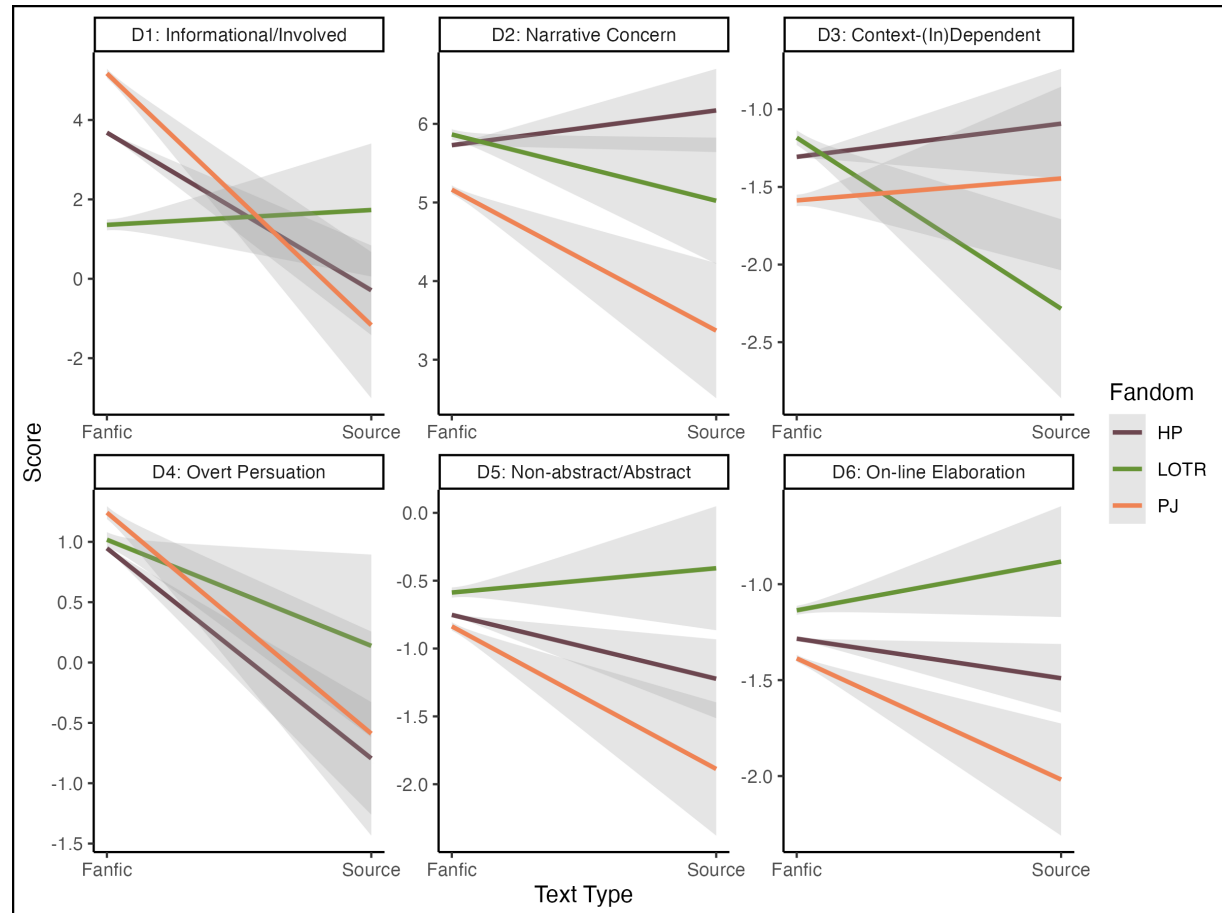AARHUS UNIVERSITY
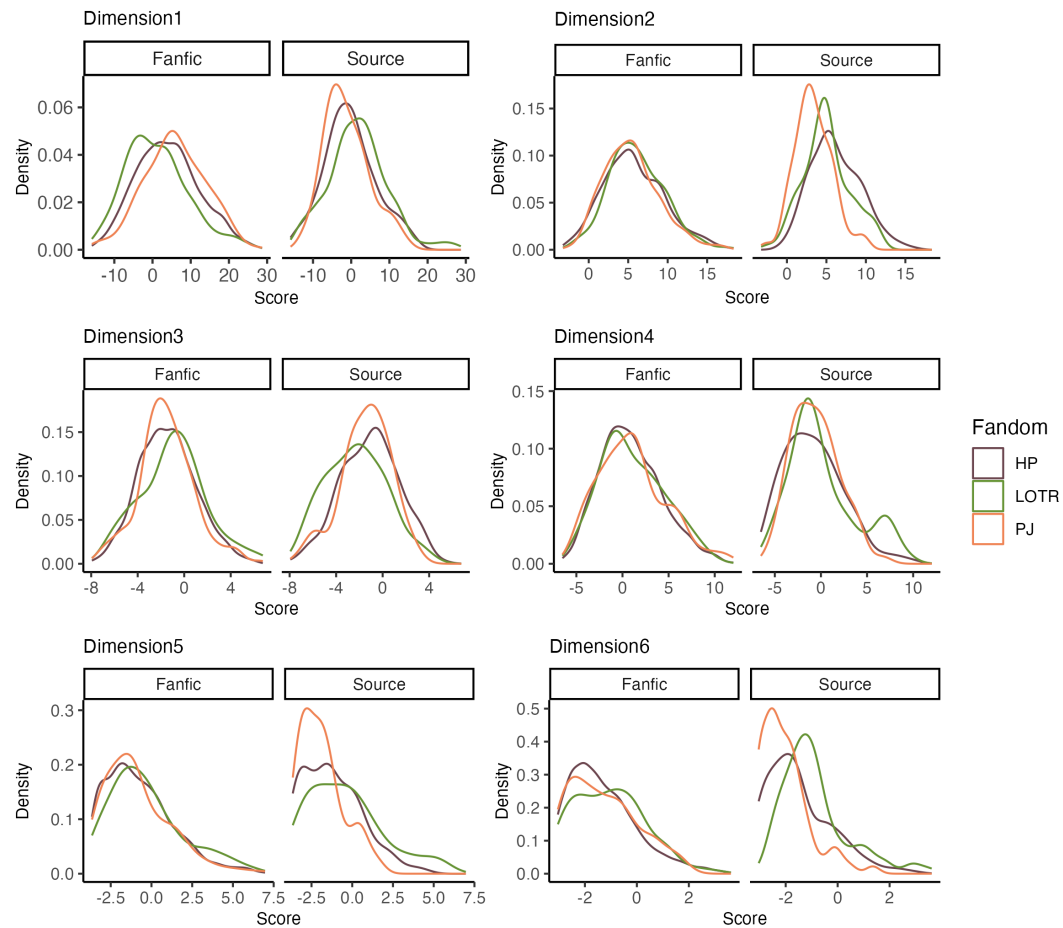
RESULTS

LOTR
- Most informationally heavy, least conversational

PJ
- Least informationally heavy, most conversational

HP
- In between the two others
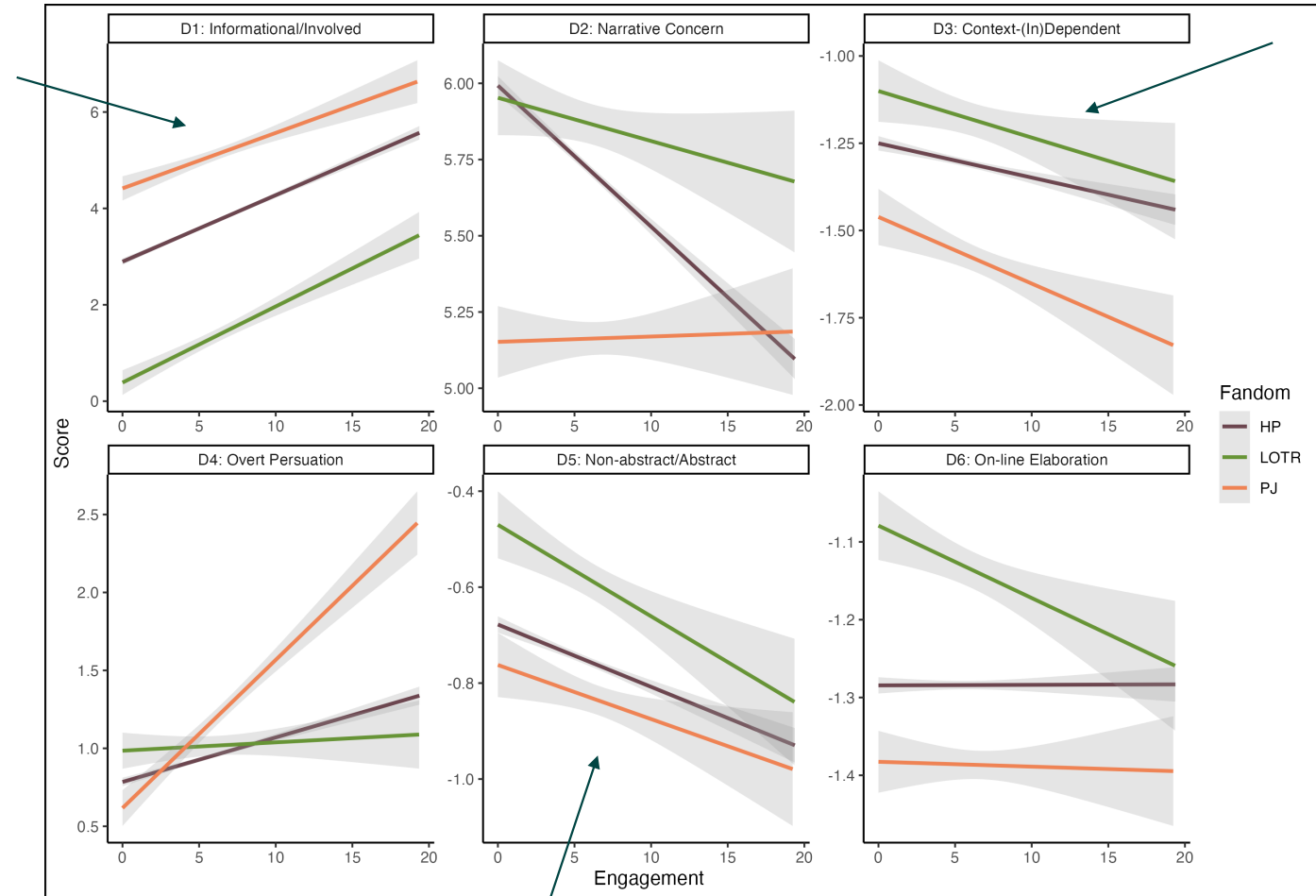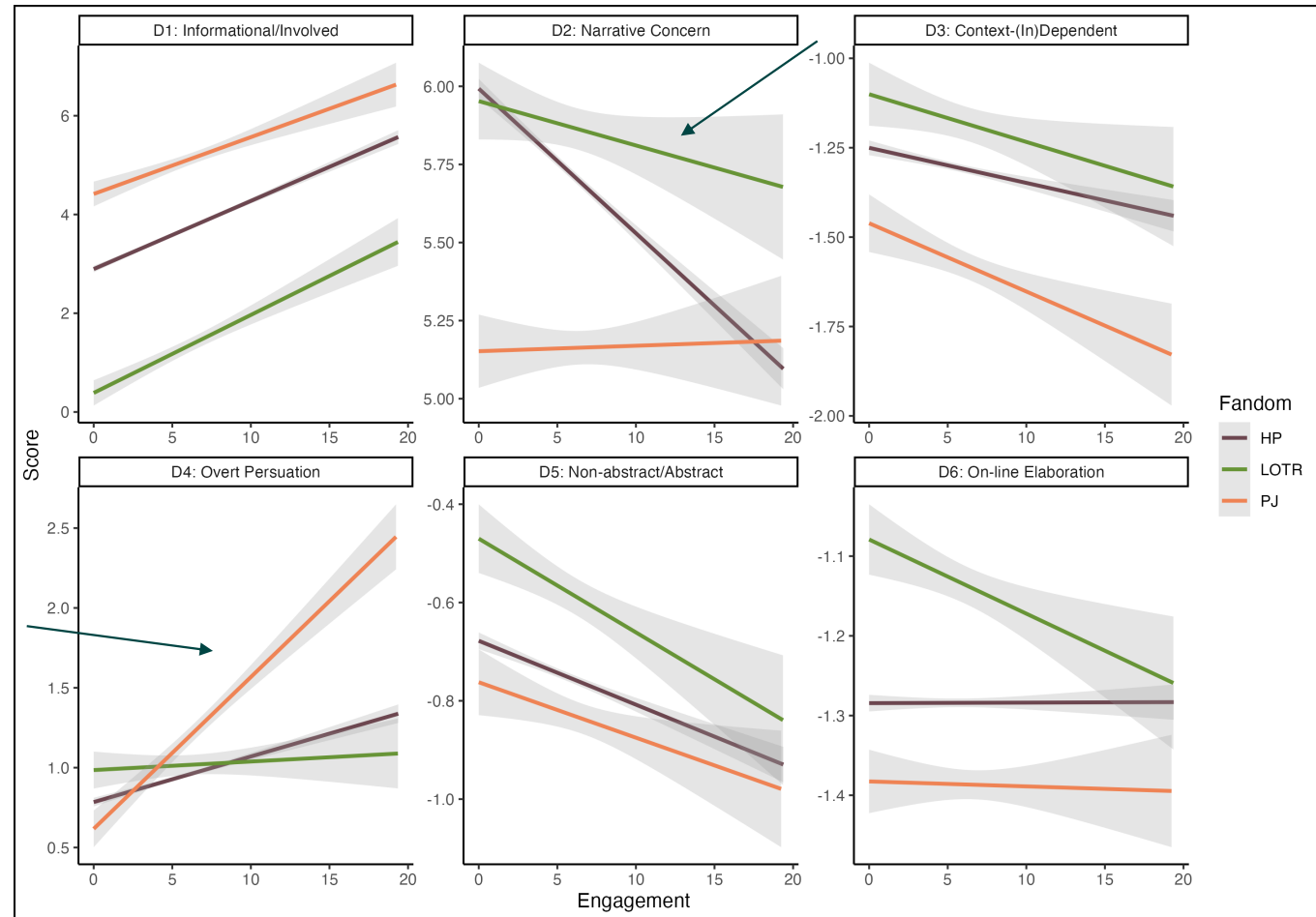
RESULTS

Fans in general prefer more conversational stories

SCHOOL OF CULTURE AND SOCIETY

AARHUS UNIVERSITY

WORKSHOP
11 JUNE 2025

MIA JACOBSEN
RESEARCH ASSISTANT

18

RESULTS

Fans show community-specific preferences for the features that set them apart from others

CONCLUSIONS

★ Fanfiction writing does not differ from the source material when it comes to Biber's functional dimensions of variation

★ There does seem to be a general fanfiction style which the style of the source is super-imposed on

★ Fans generally prefer here-and-now oriented stories that are less formal and informationally heavy

★ Specific fandoms also show greater preference for the textual traits of the source texts that set them apart from other fandoms

AARHUS
UNIVERSITY