The Stickiness of Category Labels: Audience Perception and Evaluation of Change in Creative Markets

Balázs Kovács
Yale University
School of Management
165 Whitney Ave.
New Haven, CT 06520
Balazs.kovacs@yale.edu

Greta Hsu
University of California-Davis
Graduate School of Management
540 Alumni Lane
Davis, CA 95616
grhsu@ucdavis.edu

Amanda Sharkey
Arizona State University
W.P. Carey School of Business
P.O. Box 874006
Tempe, Arizona 85287
amanda.sharkey@asu.edu

May 2023

FORTHCOMING IN *MANAGEMENT SCIENCE*

*We are grateful to Clayton Childress, Gina Dokko, Michele Piazzai, and Elizabeth Pontikes, as well as participants in the seminar audiences at the University of California-Berkeley, HEC Paris, Columbia University, Emory University, UC3M, London School of Economics, University of Oxford, Yale University, NYU Abu Dhabi, the 2021 Nagymaros conference and the 2021 Cultural industries conference, for their helpful comments and suggestions on earlier drafts of this work.

Abstract

Market producers often seek to position themselves in different categories over time. Successful repositioning is difficult, however, as audiences often devalue offerings that depart from a producer's past creations. Prior research suggests that this penalty arises as evaluators withhold opportunities for producers to reposition due to presumptions of a lack of competence in different categories. In this paper, we develop understanding of a novel evaluator-driven challenge to producers' repositioning efforts: evaluators are prone to "categorical stickiness," where the categories they have come to associate with a producer through its prior offerings shape their perceptions of the producer's subsequent offerings. The result is a systematic mismatch between what producers claim and what evaluators perceive when a producer repositions. We further propose that audience members who have the greatest prior experience with a producer will be the least likely to recognize its repositioning efforts. We examine evidence for our theory using data from Goodreads.com on authors within the book publishing industry, 2007-2017. We first build a novel deep-learning framework to predict categorization of a given book based solely on an author's description of its content. We then use data on how Goodreads users categorize and evaluate books, as well as their past reading behavior, to test for evidence of our proposed mechanism. Overall, our results extend understanding of the evaluative processes that generate categorical constraints and how these may differ among various types of audience members.

Keywords: repositioning, categorization, audiences, book publishing industry, reviews, deep learning, natural language processing

Producers seek to change their market positions over time for a variety of reasons. Increased competitive pressures, technological change, and performance shortfalls within a given market position push producers to target new positions within a market's category structure (Benner 2010, Carnabuci et al. 2015, Dobrev et al. 2001, Eggers and Song 2015, Haveman and Nonnemaker 2000, Kaplan and Tripsas 2008, Schimmer and Brauer 2012). Producers may also become emboldened from histories of competitive success to explore new categorical opportunities (Barnett and Pontikes 2008, Negro et al. 2022). Taking strategic actions to acquire and demonstrate competencies outside the bounds of their existing categorical identity may be particularly important for producers in creative industries, where the emphasis on creative expression and originality pushes producers to explore new paths (Hirsch 1972, Jones et al. 2016, Godart et al. 2023). Repositioning—changing one's categorical position over time[1]--can not only increase a creative producer's future opportunities and extend their longevity, but also speaks to the desire for creative exploration, challenge, and variety (Durand et al. 2007, Ibarra 2003, Zuckerman et al. 2003).

Yet, successful repositioning is challenging. Studies of serial entrepreneurship, for example, find that producers who branch out into new contexts face hurdles such as high uncertainty, lack of context-specific experience, capabilities, and knowledge, and inability to leverage past relationships with key stakeholders (Chatterji 2009, Eggers and Song 2015, Lahiri and Wadhwa 2021). Even when agents overcome such production-side barriers to repositioning, challenges remain on the audience side of the market interface. A fundamental audience-side challenge is that evaluators form identity expectations based on producers' historical offerings (Leung 2014) and tend to assume that distinct capabilities are required for competence in different categories. As a result, past specialization in one market category becomes interpreted as a signal of lack of competence in others. This skepticism toward producers who

[1] In our view, repositioning is related but distinct from the phenomenon of category-spanning, which can occur when a producer develops offerings that are atypical in the sense that they seem to fit multiple categories, or none at all. A producer that avoids category spanning, by creating a sequence of products that each are very typical of the category in which they are positioned, might nonetheless reposition over time by creating new products that depart from the category in which the producer had previously participated.

seek to prove themselves in a new category can lead decision-makers to withhold opportunities —a phenomenon referred to as "typecasting" (Faulkner 1983, Zuckerman et al. 2003). Due to typecasting, a record of strong affiliation with one category can become a liability when a producer seeks to branch out.

Existing research generally conceptualizes typecasting as a process whereby certain producers are not given a chance to demonstrate their ability in a new category, because of the presumption that they lack the skills and abilities to succeed. For example, seminal works on typecasting examined the film industry, where directors tend not to offer film roles to actors who have concentrated their work in a different genre (Faulkner 1983, Zuckerman et al. 2003). This focus on the pre-production limitations faced by highly specialized producers is consistent with definitions of typecasting as the "curtailment of the opportunities available to job candidates based either on social attributes such as … past work" (Zuckerman et al., 2003: 1021).

In this paper, we develop an understanding of a novel evaluator-driven challenge to producers' repositioning efforts: a systematic mismatch between what producers claim and what evaluators perceive when a producer attempts to reposition. More specifically, we propose that the beliefs evaluators have formed about the producer's identity based on their past categorical affiliations will tend to carry forward, affecting how the producer's new offerings are perceived and interpreted prior to consumption, a phenomenon we refer to as "categorical stickiness." We argue that this process sets expectations about the content of a producer's current offerings--expectations that are unlikely to be met when a producer has repositioned. These unmet expectations, in turn, result in devaluation. Whereas extant work on typecasting centers around how the presumption of incompetence forecloses opportunities *prior to* production, our arguments emphasize how the stickiness of a producer's categorical identity sets expectations for the content of the producer's subsequent work. The result is more negative evaluations of a producer's offerings--even in cases where typecasting has not precluded attempts at repositioning. In this way, we suggest that typecasting penalties may be more formidable than previously appreciated, arising not only prior to but also after production.

We expect categorical stickiness to increase when audiences have had more experience with a producer. Through their prior direct interactions, repeat audiences form strong identity-based expectations which are likely to color their future interactions with that same producer.[2] These expectations may impose greater constraint on the likelihood that an audience member will understand and accept a producer's claims to a different role or market identity. As producers attempt to reposition themselves in the market's category structure, repeat audience members—who have greater knowledge and familiarity with a market producer but also stronger categorical expectations—may find it more difficult to understand and appreciate a producer's repositioning efforts relative to a new audience member with no history of exchange with the producer.

To empirically investigate categorical stickiness and its effects on audience reactions to repositioning, we study authors within the book publishing industry. In this context, genres (e.g., science fiction, romance, and mystery) are shared categorical understandings that producers such as authors and publishers use to convey the kind of book they are offering in a way that audience members such as book readers can comprehend (Childress, 2017). We investigate the processes by which the genre labels that audiences have attached to an author's prior books shape how the author's current books are perceived and labeled, and how this in turn shapes reception to them. In contrast to much existing research on change in organizations and careers, which documents internal challenges to change (e.g., recognizing opportunities, engaging in risk-taking, developing new skills and capabilities), we examine producers who have already overcome such hurdles. They have conceptualized, written, and published books--some of which deviate substantially from their previous works in their genre-related positioning.

We analyze data from Goodreads.com, a popular website that readers use to organize and evaluate books that they have read or plan to read. Using this data, we assess Goodreads users' categorization, consumption and evaluation decisions for 518,889 books written by 168,669 authors. To

[2] We focus on category-based understandings as general beliefs about and expectations formed regarding a market producer's features and offerings. This is broader than related phenomena such as reputation, which is specific to audiences' expectations of a producer with respect to quality or level of performance (Rindova et al. 2005, Washington and Zajac 2005).

help inform our understanding of the context, we also conducted interviews with 11 book authors---a set whose work spans diverse genres and includes New York Times bestselling authors, literary award-winning writers, authors who work with small and independent publishers, as well as self-published authors--and 13 Goodreads users.

To measure an author's intended positioning of their books, we build a deep learning framework to predict the categorization of a given book based solely on its producer-supplied book synopsis[3]--the summary descriptions provided by authors and/or publishers of their books (typically found on a book publisher's website, the book's back cover, or dust jacket). These synopses give readers a preview of the book that is designed to succinctly convey what kind of book it is, including key features such as main characters, setting, and overall plot. This predicted categorization indicates the publisher's/author's perspective regarding where their book should be positioned relative to the market's shared categories. We assess audience responses to these positioning efforts by utilizing a unique feature of our context: audience members can apply genre labels to each book. We analyze the extent to which an author's existing genre-based identity (i.e., their previous categorical affiliations) affects how new versus repeat readers of that author categorize their new work, net of the predicted categorization based on the book synopsis. Prior research, which focuses on category labels claimed by producers (e.g., Hsu, Hannan, and Koçak, 2009), assigned via third-party archival sources (e.g., Ruef and Patterson, 2009), or inferred from audience coverage patterns (Zuckerman 1999) could not investigate the extent to which the labels audience members explicitly assign to a producer and/or its offerings shapes their reception in this way.

Given the empirical challenge of isolating our proposed mechanism from alternative processes demonstrated by existing research, we test our theory in a variety of ways and levels of analyses. Our main analyses examine processes and outcomes at the book-level, and include a matching strategy designed to compare books from authors with similar track records, success, and category positioning.

---

[3] We validate the deep learning model by showing that it provides good out-of-sample prediction power and also by conducting a controlled online survey with human raters, whose classification of the books correlate highly with the values predicted by the deep learning model.

We also explore our proposed mechanisms at the individual reader level, showing for example that the specific order in which a reader consumes an author's books influences their labeling of the author's future books. We also explore factors such as a reader's variety in genre preferences and the diversity of an author's portfolio of books to build support for our theory. Although drawing clear causal inferences remains a challenge given the correlational nature of our data, we believe our series of tests still extend and enrich existing understanding of audience-driven penalties for repositioning.

In the next section, we develop theory on categorical stickiness and the role it plays in constraining audience acceptance of market producers' repositioning efforts. We further consider how audiences' acceptance of a producer's efforts to switch to new genres varies depending on the extent to which audience members have prior direct experience with the producer.

Categorical heuristics in markets

We first consider how audiences are influenced by producers' categorical identities. Different categories in a market map on to different sets of features their respective members are expected to adopt. For example, hedge fund categories correspond to distinctive investment styles associated with different risks, returns, and rates of fund survival (Smith 2011). Labor market categories correspond to different types of skills, training, and areas of work (Zuckerman 1999). In the creative industries, genre categories are represented by different sets of aesthetic features or practices that artists use to position their works in their cultural fields (Becker 1982, Negro et al. 2022).

Recognizing the role of categorical expectations in markets, producers tend to position their products vis-à-vis shared categorical understandings. As part of their positioning efforts, producers may select certain product features, competitive behaviors, and pricing to signal particular positions (e.g., Porter, 1980; White, 1981; Dobrev, 2007). Producers also advance claims through strategic use of labels (Granqvist et al. 2013, Lounsbury and Glynn 2001) and carefully crafted identity narratives (Whetten, 2006; Kennedy, 2008; Ibarra and Barbulescu, 2010; Glynn and Navis, 2013).

Of course, while producers may claim a particular market position through strategic feature selection and identity narratives, a producer's positioning efforts may influence but cannot determine the identity-based beliefs and expectations that others form for it (Gioia et al. 2010a, Ravasi and Phillips 2011, Whetten 2006). Audiences interpret producers' claims to a specific position within the market's shared structure from the vantage of their own market experiences. Market audiences may accept some aspects of a producer's claims regarding its identity, while ignoring or rejecting others (Hsu and Hannan 2005). Over time, the cumulative choices that producers make regarding the positioning of their individual products -- and the way in which those products are received – result in the formation of a producer's overall identity (Gioia et al. 2010b, Glynn and Abzug 2002, Lounsbury and Glynn 2001).

Prior research suggests that the categorical identity with which a producer has become associated over time in turn shapes perception and recognition of their current work (Leung 2014, Younkin and Kashkooli 2020). In many cases, this can be beneficial, as producers with strong identities gain the benefits of recognition and can expend less effort and resources establishing the nature of their offerings to potential audience members. But it also suggests that, to the extent that a producer has been recognized as a member of a particular category in the past, audiences will presume their new offerings will continue in the same category.

Thus, while having an identity that strengthens recognition as a category member is generally beneficial for producers, it also presents constraints. When market producers seek to move to new positions, the identity audiences have associated with the producer based on prior offerings will affect their perception of the producer's new offering, causing audience members to interpret it from the vantage of the identity they already hold for the producer. For producers attempting to reposition within the market's category structure, this is likely to result in a *label mismatch*, as a producer's offerings are labeled by audiences differently than what the producer intends.

> **Hypothesis 1:** The more a producer repositions from prior offerings to the current offering, the greater the mismatch that will arise between the producer's positioning and the audience's labels of its current offering.

The extent of label mismatch is expected to increase with the audience member's prior direct experience with a market producer due to the clearer, stronger category-related beliefs they hold about a market producer's identity. For example, studying financial analysts' responses to a firm's attempt to reposition its identity from a digital photography to a flash memory company, Tripsas (2009) finds that, while all analysts were slow to update their categorizations of the firm, the single analyst who covered the firm before its repositioning efforts failed to update throughout the study period even after other analysts had switched categorizations. Similarly, Benner (2010) shows that securities analysts, who tend to have long-term relationships with the firms they cover, paid little attention to photography and wireline telephone companies' efforts to offer products that differed substantially from what they had produced in the past, focusing instead on products that were closer to the firms' historical roots. If a similar dynamic holds among evaluators more generally, label mismatch will be greater when the audience for a producer's current offering has more experience with the producer's prior offerings.

> **Hypothesis 2:** Label mismatch will be greater for audience members who have prior exchange or consumption experience with the producer.

Label mismatch, audience attention, and appeal

Category expectations often correspond to general distinctions in the needs or preferences of subsets of the market audience (Hannan et al. 2003, Hsu 2006). That is, expected category features tend to be salient markers around which audience preferences and consumption patterns cohere. Such tendencies may be particularly strong in cultural markets, where fans often develop strong loyalties to their preferred or core genres (Younkin and Kashkooli 2020).

This general correspondence between categorical expectations and audience taste preferences means that market producers who establish a clear identity that conforms to category expectations tend to attract greater audience attention. Such producers are easier for audiences to identify and compare. A clear categorical identity further generally increased perceived audience appeal, as audiences tend to appreciate offerings that meet their category-based expectations (Hsu 2006). Both attracting attention from resource-holding audiences and effectively appealing to them are important markers of the success of a producer's offering. And, in markets where word-of-mouth and online review sites influence everyday consumers' decisions, these two outcomes are likely to be related, as positively-valenced reviews can fuel greater overall audience attention and demand (Sharkey et al. 2023).

Producers who have established a category-specific identity, however, are also likely to face considerable constraint from audiences regarding their future offerings. Although a producer may have repositioned, audiences will tend to approach their new offerings through the lens of a different, prior category identity. As a result, they will expect features that the current offering does not deliver. This mismatch between audience members' ex ante expectations and what the offering delivers may lead audiences to quickly discount or drop the offering from serious consideration. Adding to this, audience members who consider such offerings are likely to find them less suited to their specific tastes, given the mapping of expectations to preferences. This increases their likelihood of disappointment and distaste. And, in markets with online review sites, this lowered perceived appeal (expressed through lower ratings) may deter other readers from considering an offering, further decreasing the overall attention an offering receives. Accordingly, when there is greater divergence between a producer's intended category position and the category labels assigned by audiences, we expect a decrease in both audience attention to and the perceived appeal of an offering.

> **Hypothesis 3**: The greater the label mismatch, the less appealing audience members will find an offering to be and the less overall attention it will receive.

The overall implication is that label mismatch will at least partially mediate the relationship between repositioning and lowered attention/appeal. While there may be other factors shaping this relationship (for example, readers may understand what the author is attempting but still feel disappointed in the product experience), we expect label mismatch to account for a portion of any negative impact repositioning may have on audience attention and appeal.

DATA AND SETTING

We test our theory in the book publishing industry. In this context, genre categories play an important role in structuring market interactions (Childress 2017). Market producers in this setting (i.e., book authors) often specialize in genres and develop identities associated with them. For example, Stephen King is a well-known author strongly associated with the horror genre due to widely read books such as *It* and *Carrie*, and Danielle Steel is strongly associated with romance due to books such as *The Gift* and *Big Girl*. Genres have common features or practices that readers have come to expect. A Gothic mystery tends to take place in an old castle or manor, with a young, innocent female as a main character, while a science-fiction novel features futuristic scientific or technological advances. An author we spoke with noted that "readers who follow a particular genre expect the story will unfold in a certain way that is consistent with the genre. If you don't do that, they'll stop reading…You've basically broken the promise to the reader." Specializing in a given genre allows authors to develop expertise and familiarity with common practices. It also facilitates publishing, since literary agents and book publishers tend to specialize by genre (Bransford 2018). And, importantly, readers often search and specialize by genre. The readers we interviewed typically read within 1-3 genres, although some noted they might try other genres at times. Authors who can craft a clear identity within a well-established genre can match with a base of readers who appreciates its common practices—a clear advantage within the crowded book publishing world.

Yet, authors branch out into different genres at times. For example, Stephen King also wrote *The Green Mile*—a prison drama with elements of fantasy. Switching genres allows authors to pursue new creative directions. As one author we interviewed noted, "you need to stretch your legs and you need to see what else you have inside you. Even if it fails." And, relative to other industries, it does not require large capital investments, making it more feasible for book authors to change genre positions. Yet, repositioning is often regarded as a risk. Writing in a new genre not only requires the acquisition of new skills and techniques, but also runs the risk of alienating loyal readers who have come to appreciate the author for a given type of story and who may be confused and upset by the new direction (Gold 2014).[4] In our empirics, we explore the role that category stickiness plays in the consequences associated with genre repositioning by authors.

We analyze data from Goodreads.com, "the world's largest site for readers" (About Goodreads). Goodreads is an online community in which users search for books, post reviews, and organize books they have read or wish to read. The data on this website allows us to track authors' publishing histories and link this to users' behaviors over time. The Goodreads data contains all publicly available data posted between the site's founding in 2007 and before December 2017[5]. The full sample contains information about 1,521,962 books written by 462,844 authors. Testing our theory requires us to restrict our sample in several ways. First, we include books that are published in English[6], published in or after 1970, and have a single author (so that we can unambiguously capture an author's history of prior work[7]). To avoid inaccurate classification, we also restrict our analyses to books that have a textual description (synopsis) that is at least ten words long, and that have received at least five crowdsourced genre tags (see below for greater detail on how tagging occurs on Goodreads). Finally, because we wish to study audience reactions to authors' genre positioning over time, in our analysis sample we only include authors who have at least two books that fit the above criteria. In our sample, authors published an average of 2.64 books; authors with at least two books published an average of 6.14 books.

Our sample represents authors who have generally achieved a higher level of success relative to the broader universe of book authors. An author we interviewed noted both the ratings counts and

[4] To avoid typecasting, some authors publish under a pseudonym. For example, JK Rowling published her Cormoran Strike books as Robert Galbraith. For a small proportion of books (0.06%), Goodreads lists a crowdsourced label for author pseudonym. It is likely that this is indicated for more well-known authors with larger fan bases. To ensure that our results are not driven by pseudonyms, in supplementary analyses (available upon request), we drop all authors where a pseudonym is indicated (e.g., both JK Rowling and Robert Galbraith are dropped) and find effects consistent with our main analyses reported.
[5] It does not contain reading and review information that the reviewers made private or only shared with their friends.
[6] We analyze the data on the book level, combining reviews for different editions of the same book. If a book has multiple editions and/or has been published in multiple languages, we take the description of the first English version.
[7] Approximately 94% of books on Goodreads.com are single-authored.

numerical ratings on Goodreads to be important metrics for judging a book's success. In our dataset, whether an author publishes a second book is more strongly related to the number, rather than the level, of ratings their first book received. A one standard deviation increase in the log count of ratings of an author's first book increases the likelihood that the author will have a second book by 13.7%. In comparison, a one standard deviation increase in the mean rating of the first book increases the likelihood of a second book by only 1.2%.

The resulting sample contains 518,889 books by 168,669 authors and 96.2 million numerical ratings from Goodreads users. For each book, Goodreads shows the synopsis that book publishers and authors have written to capture the book's main features, including key topics and themes, plots, setting, and main characters. The authors we interviewed described iterative processes through which book synopses were developed and refined in order to distill a book to its "essence." Childress (2017: 142) observes in his study of book production that editors aim "to get to clear categories" and spend a considerable amount of time figuring out how to formulate, highlight, and balance text in the synopsis in order to avoid the possible miscategorization by audiences of books into the wrong genres. Through this process, the editor, in conjunction with field representatives, marketing staff, and the author, develops a synopsis designed to place books in the hands of the "right audience." The self-published authors we interviewed described similar iterative processes with contracted editors and trusted sources such as experienced reviewers and other authors. On the other hand, the readers we interviewed typically noted first reading book synopses to determine if a book will be the kind they would like.

Book labeling at Goodreads is crowdsourced to readers. When a reader adds a book to their profile, they may choose to put the book on a bookshelf, akin to labeling it. Goodreads provides default shelf types (i.e., "read", "currently reading," and "to read"), but users are also free to choose any names for their bookshelves.[8] In our interviews, readers reported using genre tags, such as "fiction" "sci-fi" "romantic comedy" or "chick lit" as bookshelf names to sort their books. Users can put a book on

[8] We do not use the "to-read," "currently reading," or "read" tags as labels in our analyses.

multiple shelves. Taking into consideration all the labels applied to a book by different readers provides a picture of how the book is generally perceived. For example, a given book might be labelled as “fiction” by 653 users, “sci-fi” by 324 and “romantic comedy” by 149 users.

Goodreads users have used thousands of unique shelf names. Many shelf names are rarely used and the distribution is long-tailed. Because computation limits make predicting labeling vectors with so many cells infeasible and because we would need a larger dataset to train our deep learning model, we limit the set of labels to those 38 higher-level genre classifications listed on Goodreads.com (designed for user browsing; see Table A1 for the list of genre labels). These crowdsourced labels allow us to capture directly how Goodreads users perceived and categorized books in our sample. Because genre classifications may be unreliable in books with very few tags, we only include in our sample books that were tagged by readers at least five times (62.2% of books in our sample).

The genre-classification data we discussed above is publicly available at the book level in its entirety. A subset of reviewers made this information public at the individual level. For these cases, we could also download the genre labels applied by each individual reader. For example, we may know that “Erica C” tagged a book “sci-fi”, “Michael T.” tagged it as “mystery”, while “Apple D.” tagged it as “mystery” and “sci-fi.” On average, readers in the individual level dataset assigned 1.34 genre labels to each book. In total, this dataset consists of 2 million individual genre labels assigned to 253,161 books. Given that this is not likely a representative sample, we use this reader-level tagging data to conduct supplementary individual-level tests of the hypotheses and mechanisms.

Using deep learning to separate the category stickiness effect from the book synopsis

We next turn to explaining a core facet of our empirical approach: the use of deep learning models to predict how a book is likely to be labeled based on a book’s synopsis alone. The approach we use to predict each book’s genre categorization--deep learning--is a part of a family of supervised machine learning (ML) methods that uses neural networks to perform various tasks, such as text classification. Computer scientists have developed a variety of deep learning techniques designed to understand and represent human language. The approach we ultimately selected centers on Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2019)—a major advancement in natural language processing (NLP) developed by Google researchers to improve understanding of search queries, which are often complex and conversational in nature (Nayak 2019). BERT is available as an open-source library that has been pre-trained on a vast set of text data from Wikipedia (~2,500 million words) and BooksCorpus (~800 million words) . It has been shown across multiple datasets to significantly outperform older text analysis and machine learning approaches. In the same setting that our study focuses on (i.e., the book industry), Le Mens et al. (2023) show that BERT substantially outperforms commonly used NLP techniques such as Bag-of-words and GLOVE embeddings in predicting a book's genre labels from its synopsis.

BERT is designed to overcome one of the major challenges to effective natural language representation—insufficient training data. A key intuition underlying this approach is that context—the words that come both before and after a focal word--is key to understanding each word's meaning. For example, the word "bank" can take on multiple meanings, depending on the words surrounding it (e.g., "a financial bank", "bank of a river", a "bank shot"). BERT learns to predict meanings through Masked Language Modeling, in which roughly 15% of the words in a training sample's sentences are replaced with a masked token, and the rest of the non-masked words in the sentence used to learn prediction of the masked word (Devlin et al. 2019). The result is a robust general-purpose language representation model. We fine-tune this model to our specific NLP task: predicting book genres based on books' synopses.

---- INSERT TABLE 1 HERE ----

We enter into our model each book's synopsis, truncated at 150 words to standardize across descriptions.[9] Output for each book is a 1 X 38 vector that contains the predicted weights for each of the

[9] The median book description length is 127 words, and 63% of descriptions are shorter than 150 words. We decided to truncate at 150 words because truncating at 100 significantly lowered out-of-sample prediction accuracy (by 1.5%), while including the first 200 words only increased prediction accuracy by 0.1%. We decided against using truncation at 200 words because this would introduce significant variation in the amount of information used to predict the location of various books (i.e., the location of books with longer descriptions will be more precise than

38 labels. Table 1 illustrates the data structure for a book in our sample, James Siegel's *Derailed.*[10] The table shows *Derailed's* synopsis and the 38 labels predicted by the deep learning algorithm. The "observed" column lists labels submitted by Goodreads users, in percentages of all labels assigned to the book. The "predicted" column contains the predicted label percentages based on the book's synopsis (more on this below). Both "observed" and "predicted" columns were normalized such that they sum to 1. One can see in Table 1, for example, that 12.9% of the labels users applied to this book were "mystery," while its predicted label percentage for mystery was 19.3%.

To generate predicted label percentages based on book synopses, each text is first "tokenized," meaning that the text is turned into numbers that represent the text. To make the vectors for the books comparable, each book is represented by a 1 X 150 vector. If a synopsis is shorter than 150 words, it is right padded with zeros. This makes the vectors representing texts comparable and is consistent with common practice in machine learning (Chollet 2017). We do not remove punctuation and stop words, as these are important part of the text representation in the BERT model (Devlin et al., 2019).

We next feed the vectorized data into a deep learning network. The network structure we use is built from a 150-dimensional word embedding layer, 12 BERT layers of 768-dimensions, a 38-dimensional fully connected layer, and a 38-dimensional softmax layer which outputs the predicted categorization probabilities[11]. We then fine-tune (i.e., train) this representation for the Goodreads label data. The training model minimizes the categorical cross-entropy between the predicted and the observed categorization vectors.[12]

---

that of books with shorter descriptions). The book locations estimated based on text truncated at 150 or 200 words correlate very highly (>98%).

[10] For the book's profile on Goodreads, see https://www.goodreads.com/book/show/314393.

[11] We use the BERT-base-cased representation, which itself consists of 12-layers, 768-hidden dimensions, 12-heads, and 110M parameters (see https://github.com/google-research/bert). The softmax function is a generalization of the logistic function to multiple dimensions. It is used in multinomial logistic regression to normalize the output of a network to a probability distribution over predicted output classes (Goodfellow et al. 2016).

[12] The deep learning algorithm was programmed in Python using the Keras package and trained using Google's Colab cloud services. The code is available from the authors.

The dataset used to train our model should be similar to but distinct from the main dataset. Since our theoretical arguments around repositioning dictate that our analyses focus on authors who have published at least two books, we use books written by authors with only one book (114k books) as our training sample. Because an author publishing their first book has no history of prior genre affiliations that might influence user labeling, we assume that labels of authors' first books are generally more likely to correspond to the books' synopses. The second set of books (those written by authors with at least two books (404k books)) is the out-of-sample prediction sample. Our regression analyses are all based on these out-of-sample predictions.

The second consideration in constructing the training sample is the temporal nature of our data. Specifically, we want to train the model such that the training set of books is close in time to the predicted set of books because the meaning of genre classifications may change over time (i.e., what people meant by "romance" in 1970 may be different from what people mean by "romance" in 2015). Moreover, we want to avoid using future data to predict past behavior. We achieve this dual goal by constructing a 10-year moving window sample for training the model for each cohort. For example, to predict the genre classification of books published in 2010, we train the model on the subset of books in the training sample that were published between 2000 and 2009 (inclusive). The more we go back in time, the fewer books there are in the training sample. Therefore, we decided to limit the out-of-sample set of books to those published in 1970 or later and to authors whose first book was published in 1970 or later. Then, for each year, we estimate the model on the prior 10 years, and calculate the out-of-sample predictions. The model is trained for one epoch[13] as we found that one epoch provides the best model fit without overfitting.

Validating the deep learning model with human raters

We assessed how well our deep learning model performs in two ways: (i) examining out-of-sample prediction performance, and (ii) by collecting human classification in a controlled environment

[13] An "epoch" in machine learning indicates the number of passes of the entire training dataset the learning algorithm has completed (see e.g., https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch)

and showing that the model predictions are not biased by the author's level of repositioning or the book's publication order.

First, for out-of-sample model fit, we assess the correlation between predicted and observed labels. For the training sample of authors' first books, this correlation is 78%. Figure 1 plots the relationship between observed and predicted label proportions, averaged across all book-label pairs in the out-of-sample prediction set. For example, the point (0.65, 0.60) on the figure indicates that, on average across all books and labels, when 60% of a book's observed labels are in a given category (e.g., romance), our deep learning model predicts that 65% of the book's labels would be in that same category. If the observed values were the same as the predicted values, this figure would show a 45-degree diagonal line. The relatively linear and monotonic pattern we see in Figure 1 indicates an overall good model fit.

---- INSERT FIGURE 1 HERE ----

Second, we examined how well the predictions of our deep learning model corresponded to human labeling in a controlled environment. To do so, we conducted a study on the online platform *Prolific* in which human raters were asked to read book synopses and tag them using the 38 Goodreads genre labels (see Appendix Figure A1 for an example of the task). We randomly selected 125 books from our out-of-sample prediction set such that 1) they were published after 2010, 2) the book synopsis was between 500-700 characters long, and 3) the book was the 1st, 2nd, …, 5th of books published by its author (we randomly selected 25 books for each bracket). We recruited participants who were U.S.-based, listed English as their first language, and had at least a high-school education. We recruited 700 participants, 640 of whom passed an attention check. We use the results of those who passed the attention check.

In the experiment, we presented participants with 5 randomly selected books in a randomized order. On average, each book was tagged by 640*5/125=25 people. For each book, we aggregate the individual genre tags, so, for example, a book may have been tagged as mystery by 8 people and romance by 12. We then normalize the tag counts by book such that the label proportions sum to 1 (i.e., 40% mystery and 60% romance). Thus, each book's human ratings are represented with a 1 X 38 vector, in which the values are the label proportions for that specific genre (0 for those genres that were not applied by anyone). We compare these values to the label proportions predicted by our deep learning model. Overall, we find a 73% correlation between the survey-based genre labels and the prediction of the deep learning model. Because the book synopses are shown in the survey without their author's name and without any information on the previous books of the authors, there is no reason to believe that the human ratings are biased by knowledge of the author's prior books. Moreover, we ran regressions in which the dependent variable was the difference between the survey and the deep learning predictions and the independent variables were the extent of repositioning and publication order. None of these variables had a statistically significant effect. Overall, these results suggest that the deep learning values are not biased systematically by publication order and extent of repositioning.

In addition to validating the performance of the deep learning model, we also considered how the predicted label proportions outputted by the deep learning model relate to and capture several key properties of label spaces identified in the categorization literature, including distances between labels, differing levels of abstraction, and category fuzziness. See Appendix B for discussion of these issues.

Variable construction

Outcome variables. Our theoretical arguments center on three main outcomes: the extent of mismatch between the labels assigned by readers and the labels predicted based on the producer's synopsis (label mismatch), audience attention, and audience appeal. We discuss each in turn.

Label mismatch. We calculate the mismatch between the labels assigned by readers and the labels predicted based on the producer's synopsis, by calculating the mean squared distance between the observed label vector and the predicted label vector. Specifically, for book i,

$$\text{Mismatch(i)}=\sum_{d=1}^{38}[observed\ label(i,d) - pred.label(i,d)]^2$$

Audience attention. To measure attention, we examined the count of users who rate a book. While readers who review a book may not necessarily like the book, they typically will have enough awareness and familiarity with the book to enable them to evaluate it. In 2007 (the first full year Goodreads was in operation), there were 13,053 unique users who left ratings. In 2016 (the last full year in our sample), 426,083 unique users left ratings.

Audience appeal. To capture the appeal that a book has generated, we examine the numerical ratings assigned by users to books (1-5 integer star rating). The mean star rating assigned is 4.09. 2% of the ratings are one star, 6% are two stars, 22% three stars, 36% four stars, and 34% five stars. In our book-level analyses, we examine both the mean rating and the count of 5-star ratings, since this latter measure distinguishes books with a relatively rare level of positive appeal.

Covariates.

Extent of repositioning. We use book synopses to calculate the extent to which an author attempts to reposition. We first calculate each book's predicted position in label space based on its synopsis. We then calculate the distance between the focal book's position and each of the author's previous book(s)' positions. Specifically, we measure the sum of squared distance of the focal book's synopsis from that of each previous book in the predicted label space and average these pairwise distances. Formally, we use the following equation:

$$distance(focalbook_i \ , from\ previous\ books) = \frac{\sum_{j \in previous-books} \sum_{d=1}^{38} [pred.label(i,d) - pred.label(j,d)]^2}{|j|}$$

where j denotes the set of previous books by the focal author, and d is a counter from 1 to 38 for the predicted label dimensions.

Appendix Figure A2 shows the distribution of these distance values. As the figure shows, most books are highly similar to previous books by the author, (i.e., most authors do not change very much), but there are also a large number of authors who reposition substantially.

Repeat readership. To measure the proportion of a focal book's readers who have had prior experience with an author, we examine whether each reader has completed a book by that same author prior to the focal book. To be counted, books read earlier do not necessarily have to be books published earlier—our measure focuses on the timing of a user's prior interaction with the author rather than the author's publication timing. Approximately 82% of the interactions are from users completely new to the author. Roughly 61% of readers who read one book by an author go on to read a second book by the same author. The average Goodreads user has read 101 books by 57 unique authors.

In our book-level models, we account differences across authors by either matching on or controlling for several variables. These include the logged count of books the author had published prior to the focal book (*author book order, ln*), *the mean rating of an author's prior books (*weighted by the count of ratings for each book)*,* the *publication year of the focal book* and, in models estimating label mismatch, the logged number of labels the book had received (*total labels applied to the focal book, ln*).

Analytical Approach

Our main analyses focus on book-level dynamics. For models estimating label mismatch (testing H1 and H2), our main models focus on authors with at least two books, and we use author fixed effects (N= 308,795 books). For models predicting attention and appeal (H3), we construct a matched sample of books (N=262,503 books). We conduct supplementary analyses at the individual-reader level (N=1,285,992 ratings) to better isolate underlying mechanisms as well as post hoc analyses that explore theorized relationships further. Table 2 provides an overview of the different tests we conduct, while Tables 3a-3b provide descriptive statistics and pairwise correlations for variables used in the main (i.e., book-level) models and Appendix Table A2 provides descriptive statistics for the supplementary models.

---- INSERT TABLE 2 & TABLE 3 HERE ----

For our first set of models (testing Hypotheses 1 and 2), the dependent variable in our book-level models is label mismatch, and main covariates are the author's extent of repositioning and repeat readership. We include author fixed effects with standard errors clustered on the authors. The sample for these models consists of the second and later books from all authors who have written two or more books. The equation for these models is:

$$Y_i=\beta_1 X_{1,i}+\cdots+\beta_k X_{k,i}+\alpha_j+u_i$$

where i refers to a book from author j, and the numbers 1 through k designate different independent variables. The α_j are group-specific intercepts that capture heterogeneity across groups (i.e., author fixed effects). Fixed effects do not completely eliminate the within-group correlation of the residuals (Arellano 1987), so we cluster all standard errors at the author level.

In Hypothesis 3 we proposed that, the greater the label mismatch, the less attention audiences will give to the offering and the less appealing they will find it. Given heterogeneity in factors that may shape a book's appeal, we employ a matching strategy using coarsened exact matching in Stata to construct the main sample for this analysis. More specifically, starting with the sample of all books published by authors who have two books (and that meet the other criteria outlined earlier), we construct a control sample by creating matched strata of books in which books with a high (i.e., above-median) degree of label mismatch are matched with another books that has low (i.e., below-median) mismatch but is otherwise observationally similar in terms of the following metrics: deciles of the total count of ratings of all prior books of the author, deciles of mean rating of all prior books of the author, and deciles of the focal book's publication year.[14] We also match (exact match) on the count of prior books published by the author, and the most common predicted genre label of the focal book. We use the CEM command of Stata 17 for this matching, which creates m:n matches with weights (see Blackwell et al. (2009) for details of the matching algorithm). Appendix Table A3 reports statistics showing minimal differences across the treatment and control samples after matching (Cohen's D < 0.15). The book-level dependent variables are the (logged) total count of users who rated each book, the (logged) count of five-star reviews, and the mean rating, and the main covariate is label mismatch. We estimate Ordinary Least Squares regressions of the following form:

$$Y_i=\beta_1 X_{1,i}+\cdots+\beta_k X_{k,i}+ u_i$$

where i refers to a book from author j, and the numbers 1 through k designate different independent variables. We cluster standard errors at the matched strata level.

We supplement each of these book-level analyses with analyses at the individual reader level, which allow us to drill down into factors such as the order in which users read an author's books to better

---

[14] For the mean rating of an author's prior books, we set this value to the population average for the first book published by an author.

understand the underlying mechanisms driving the book-level dynamics. In most of the individual-level analyses, we employ a unique matching approach: we find readers who have read two or more books from a focal author and match pairs of readers for whom the second book they read by that author was the same book. For example, if an author published books A, B, C, and D, then a reader who read A and C will be matched with another reader who read B and C. See Figure 2 for an example illustrating the set-up of this comparison using a pair of hypothetical readers (Kristine and Mark), who have each read a different first book by the author Dave, but the same second book (Book C). We then examine how differences in the distance between the first and second book each reader has read by that author creates variation in the way they individually label and rate that same second book. Individual-level models include fixed effects by books and by reader. By focusing on pairs of readers who read the same second book, we rule out the possibility that any differences in book quality for authors who reposition substantially versus those who do not are driving the results. This individual-level data is only available for a subset of readers, since not all readers make their labels public. The equation for these models is:

$$Y_r=\beta_1 X_{1,r}+\cdots+\beta_k X_{k,r}+ \alpha_r + \alpha_i +u_i$$

where r indexes a specific reader, and the numbers 1 through k designate different independent variables. The α_r and α_i are group-specific intercepts that capture heterogeneity across readers and books, respectively (i.e., reader and book fixed effects). We cluster all standard errors at the book and the reader level.

---- INSERT FIGURE 2 HERE ----

Examining the assumption that past category labels will shape the labeling of current offerings. Before conducting our main analyses, we explore whether there is evidence supporting our assumption that the category labels assigned to an author's prior books tend to carry forward and be applied to their current books. To do this, we control for the current book's content to better isolate the effects of categorical stickiness. For example, if an author's new book is frequently labelled as "sci-fi" and their previous book was also often labelled as "sci-fi," this is not necessarily evidence of category stickiness. It may simply be that the author's books both contain futuristic themes centered around technology and are positioned through their synopses in “sci-fi”. After controlling for synopsis content, we can examine the separate effect of the author’s prior genre affiliations on readers’ label assignments.

We ran models at the book-label level of analysis in which the dependent variable is the proportion of a given label in the label vector of the book (0-1 range). There are 338,121 books in the sample of authors’ later books, and each book is represented by 38 genre labels (leading to 338,121 x 38 =12,848,598 book-label level observations). As key covariates, we include the predicted proportion of a given label based on the current book’s synopsis (i.e., the current book’s “content”) and the observed proportion of the label assigned by readers to prior books by the author, averaged across all prior books (i.e., prior books’ category label assignments). We control for the total labels applied to the book (logged), publication year, the cumulative rating count of the author’s prior books (logged) and author book order (logged). We include genre label fixed effects to control for overall popularity of a label and other fixed label characteristics. We also include author fixed effects to account for unobserved heterogeneity at the author level (e.g., author quality) and cluster standard errors on both authors and labels.

Appendix Table A4 shows the results of book-label level models predicting the proportion of a given label in the observed label vector of each book. Model 1 shows that labels predicted based on the producer’s synopsis (β=0.483, $p<0.001$) and labels attached to the author’s previous books (β=0.599, $p<0.001$) both significantly shape readers’ labeling of the current book. This is consistent with the assumption that labels assigned to a producer’s prior offerings are likely to be assigned to current offerings, controlling for the content in the book’s synopsis. See the Appendix Table A5 for a further confirmation of this assumption using individual reader-level data.

RESULTS

Our first set of main analyses are presented in Table 4. Model 1 shows the results testing for label mismatch at the book level, with author fixed effects. In support of H1, we find that the larger the repositioning, the greater the label mismatch. According to the estimates of Model 1, a one standard deviation increase in repositioning leads to a 10% standard deviation increase in label mismatch (calculated as 0.214*.0877/.196).

Model 1 specifies a linear relationship between repositioning and label mismatch, but it is possible that label mismatch is reduced when an author spans very distant labels. To explore potential non-monotonicity in this relationship, we re-estimated Model 1 of Table 4 using dummies for each 5th percentile of the variable "distance from an author's previous books" (instead of a continuous distance from an author's previous books variable). Results, shown in Figure 3, indicate the effect of repositioning on label mismatch is positive and monotonically increasing, until the very high end where we see a small decrease. This shows that negative impact of repositioning marginally decreases at very distant labels.

---- INSERT FIGURE 3 AND TABLE 4 HERE ----

In Appendix Table A6, we examine whether support for Hypotheses 1 holds at the individual reader level. Whereas our main book-level models examine label mismatch as a function of an author's repositioning from their last book, these individual-level models analyze mismatch by comparing the focal book to the other books the person has already read by the author (*individual-level extent of repositioning*). By specifying this at the individual level, we can directly capture how the first book an individual read by an author impacts their labeling of the next book read by that author—a more precise test of H1.

The logic of Hypothesis 1 suggests that label mismatch at the individual-level—the mismatch between the labels assigned by the individual reader and the labels predicted based on the producer's synopsis--will be greater when the previous book an individual read from the same author is more different from the current book. In Models 1 and 2, we analyze all cases in which a reader has labeled at least two books of a given author and include book and reader fixed effects. Models 3 and 4 show the results of a more restricted test in which we identified pairs of readers who had read the same second book by an author but each read a different first book by that author (see Figure 2). We find 64,267 such reader-dyads. In support of H1, we find that individual-level label mismatch increases significantly with greater individual-level extent of repositioning in each of the samples we analyzed.

We explore the role readers' expectations for an author may play in patterns uncovered by examining whether readers react less negatively to an author's repositioning efforts if the author has already published a diverse portfolio of books. Several of the readers we interviewed noted that some authors have a reputation for experimenting in different genres. As readers, they were cognizant when choosing a new book from these authors that the book might be different from past offerings. While we did not explicitly hypothesize this, our theory would suggest that there should be weaker expectations, and thus less label mismatch, in such cases. We created a continuous measure, *author's diversity in prior books*. For each book, we took the set of books published by the author prior to the focal book. We then calculated, for each of the 38 labels, the standard deviation of predicted label values and then took the average of these standard deviations. This measure reflects the extent to which the author has published a diverse set of prior books. Because an author's diversity in prior books can only be calculated from the author's third book onwards, we ran book-level models on the subset of books that are the third or later book by the focal author. Results shown in Appendix Table A7 indicate a negative interaction effect between the author's diversity in prior books and the extent of repositioning on label mismatch ($p<0.001$). This suggests that the impact of repositioning on label mismatch is reduced if the author has a more diverse portfolio of books.

Repeat audiences and extent of label mismatch

We next test Hypothesis 2, which predicts that the effect of repositioning on label mismatch is greater when a greater proportion of the audience has had prior exchange experience with the producer (see Table 4). Specifically, we estimate book-level models in which the dependent variable is label mismatch and the main independent variables are repeat readership and the interaction of repeat readership with author's extent of repositioning. In line with Hypothesis 2, we find a significant interaction: the greater the repeat readership, the more repositioning leads to label mismatch (Model 3: $\beta=0.0279$, $p<0.001$).

Label mismatch, attention, and audience appeal

Our analyses thus far show that repositioning leads to greater label mismatch for an author's current works—an effect that is exacerbated when the author has more repeat readers. We next examine the implications of label mismatch on attention and appeal using a matched sample of books, which helps to account for possible unobserved heterogeneity in factors that may shape a book's appeal. Our matching strategy requires a binary measure of mismatch. Accordingly, we ran regressions that split label mismatch at the median and predict appeal as a function of this binary indicator of label mismatch.

Table 5 shows the results of analyses examining the relationship between having a high level of label mismatch and the total count of users who rate a book (Models 1-3), the count of five-star ratings a book receives (Models 4-6), as well as users' mean ratings of the book (Models 7-9). We find that a high level of label mismatch is associated with significantly lower total readership.[15] According to the estimates in Model 1, high label mismatch is associated with a 17.1% decrease in the number of readers rating the book. Similarly, high label mismatch corresponds to a 18% decline in the count of five-star ratings (Model 4) and a reduction of 0.0362 in ratings (8.0% of a standard deviation) (Model 7). We note that our main results are robust to alternative specifications, such as using a continuous measure of label mismatch (Appendix Table A8) or splitting label mismatch into terciles and comparing the top versus bottom terciles of label mismatch (Appendix Table A9).

According to Table 5, the extent of repositioning has a significant effect on the count of users who rate a book and the count of five-star ratings a book receives. We do not observe a significant effect of repositioning on a book's mean rating. To examine the extent to which label mismatch mediates the effect of extent of repositioning on count of users who rate a book and count of five-star ratings, we run Hayes' (2009) mediation model with 1,000 iterations using the continuous specifications in Table A8. Based on these estimates, we find that label mismatch mediates 16.4% of the effect of repositioning on count of ratings (ln) ($p<0.001$) and 17.3% of the effect on count of five star ratings (ln). We thus find

[15] Results (see Table A12 in Appendix) estimating counts with negative binomial models are consistent with those reported here.

partial mediation of label mismatch for these two key outcomes, but do not find mediation for mean ratings.

---- INSERT TABLE 5 HERE ----

We explore whether label mismatch affects appeal similarly among readers who are new to an author versus those who have previously read a book by the focal author. In Appendix Table A10, we split ratings counts and mean ratings by new versus repeat readers. These results suggest that label mismatch does not significantly affect the number of new readers (Model 1). However, it significantly decreases the size of repeat readership (Model 2), suggesting it may impact an author's ability to retain loyal readers. But for both new and repeat audiences, label mismatch decreases the number of five-star ratings (Models 3-4) and the average rating (Models 5-6). The impact of label mismatch on five-star ratings and the average rating is somewhat larger for repeat readers than for new readers, according to formal tests of coefficient equality ($p<0.01$).

---- INSERT TABLE 6 HERE ----

In supplementary analyses (Table 6), we examine whether a negative relationship between label mismatch and appeal as represented through user ratings holds at the individual reader level. This allows us to estimate fixed effects for the reader as well as the book, and thus control for unobserved heterogeneity that may shape the ratings assigned. More specifically, book-level fixed effects used in this model control for the underlying content, ensuring that any effects of mislabeling do not stem from fixed book-level characteristics, and individual-level fixed effects control for a reader's tastes and tendency to give more positive or negative reviews. Moreover, the results here are less reliant on the accuracy of our deep learning model, because those predictions are held constant within a book. Models 1-4 examine ratings for our full sample (as described earlier), and Models 5-7 are run on the matched sample of users who had read the same book from an author second, after reading different books from that author first.

In Table 6, we estimate the rating assigned by an individual reader to a book as a function of the degree of reader-level label mismatch. Estimates in Models 1 and 2 indicate that greater label mismatch at the individual level for a book results in a significantly lower rating from that reader. Results in Model 3 show that the effect of individual-level mismatch on ratings holds even after controlling for individual-level extent of repositioning.

In Model 4, we explore the role that individual readers' tastes may play in the lower rating assigned to books with greater label mismatch at the individual level. We expect that a greater individual-level label mismatch increases the extent to which readers misunderstand whether a focal book is the kind of book that they typically like, resulting in a worse fit between the reader's tastes and the book's content. To proxy the fit between a reader's tastes and the book's content, we calculated an individual-level predicted rating of a book by a reader. To do so, we first estimate a regression using the prior books read by each individual reader, in which the dependent variable is the rating and the covariates are the 38 predicted label weights from the deep learning algorithm. After estimating the regression, we use the "predict" command in Stata v17.0 to generate an out-of-sample prediction for the individual's rating of the focal book. In other words, we generate a measure of how much the individual is predicted to like the book, based on its genres and the genres they have liked (and disliked) in the past.

We then ran analyses in which the dependent variable is the rating of the focal book, and the main covariates are the predicted rating (i.e., our proxy for user tastes, calculated as described above using data on prior books the person has rated) and the label mismatch for the focal book. We find that individual-level label mismatch continues to lead to lower ratings (β=-0.0373, $p<0.001$), even after controlling for the taste mismatch mechanism. Second, we find that this taste mismatch mechanism partially mediates the effect of label mismatch on ratings: comparing Models 1 and 4 in Table 6, we find that ~28.30% of the effect is mediated. Still, label mismatch has a significant effect on individual ratings even accounting for this alternative mechanism.

Models 5-7 in Table 6 are estimated on the more restricted set of pairs of individuals who read the same second book from an author, but different first books (see Figure 2 again for an example). We examine how each reader in a pair differs in how they evaluate the same book, depending on their individual-level degree of label mismatch. The effects of individual-level label mismatch in this restricted sample are weaker (p-values range from 0.01 to 0.06), but consistent with the effects found in Models 1-4.

If the negative effects of label mismatch are due in part to the mismatch between a book and a reader's tastes, we should further find that readers who exhibit cultural omnivorism—greater breadth in taste profiles (e.g., Peterson and Kern, 1996; Goldberg, Hannan, and Kovács, 2016)--exhibit less aversion to label mismatch and repositioning than readers with a narrower taste profile. Omnivores tend to be more willing to cross genre boundaries in their cultural consumption. While they may be surprised by a book that is positioned differently than they expected, they are likely to still find the book more appealing relative to a reader with a narrow, more well-defined taste profile.

To explore this, we created two measures of reader omnivorousness. The first is time-varying: for each reader, we generated a set of predicted label value vectors based on books' synopses for all books read by that reader prior to reading the focal book. We then calculated, for each of the 38 labels, the standard deviation of predicted label values. Finally, we took the average of these standard deviations. In the second version, we create a reader/author specific measure, where we assess the heterogeneity of the books read by the focal reader from the focal author prior to reading the focal book. We then added these two omnivorousness measures to models estimating a reader's rating for a book and interacted the omnivore measures with label mismatch and repositioning. We ran the models on the subset of readers with at least 5 ratings in order to generate meaningful omnivorism measures.

As shown in Appendix Table A11, we find a positive interaction effect between label mismatch and reader omnivorousness on ratings (β=.463, $p<0.001$ for the omnivore measure based on all books the reader read prior to the focal book; β=.957, $p<0.001$ for the omnivore measure based on the focal author's books only). This supports our expectation that label mismatch has a significantly weaker negative effect on a book's appeal to an omnivorous reader relative to a reader with a narrow taste profile.

Overall, both the book-level and individual-level results show the detrimental impacts of label mismatch on the attention and appeal that a book receives. Results suggest that the negative association between label mismatch and these markers of success is stronger for repeat readers than for new readers. In supplementary analyses, we find that the effect is partially mediated by taste mismatch, and that the negative effects of mismatch are weaker for omnivorous readers, as we would expect.

DISCUSSION

Our findings enrich understanding of the factors that lead to constraint on category positioning in market contexts. While prior work has demonstrated penalties for repositioning, studies have largely focused on signaling or skill-based explanations for why audience members devalue repositioned offerings. In contrast, we find evidence that audience labeling, which is influenced by prior categorical identities, is a discrete mechanism that contributes to these kinds of categorical constraints on career progressions. And while existing studies suggest the processes that give rise to categorical constraints largely apply when evaluators consider market producers who are unfamiliar or new to them (Faulkner 1983, O'Mahony and Bechky 2006, Zuckerman et al. 2003), our study suggests why category constraint continues to apply after evaluators have interacted with and formed clear identities for producers.

In our context, when authors attempt to branch out and position a new book that is different from their previous offerings, readers tend to perceive and label the new book using the prior genre labels. The resultant label mismatch varies in size depending on the mix of evaluators. Label mismatch becomes more pronounced when a greater proportion of a book's readers have also read an author's prior books, presumably because repeat readers have stronger expectations regarding an author's identity.

Further, our research suggests that label mismatch negatively affects the success of authors who have repositioned, resulting in decreased audience size and lower ratings. However, while we find evidence of partial mediation of label mismatch on two of our key outcomes (total count of ratings and count of five-star ratings), we do not find evidence for mean rating. Our use of a matched sample of books by authors whose prior works were similar in ratings and popularity helps to minimize the impacts of heterogeneous factors that might affect a book's appeal. Moreover, it seems unlikely that underlying quality differences that one might suspect are associated with the choice to reposition account for our results. Descriptive statistics suggest that prior ratings play a relatively small role in this decision, although popularity seems to be more important.

Contributions

Our study contributes to work on market categorization, which considers how categories mediate market experiences and systematically impact key processes and outcomes. Studies find that the way producers position themselves vis-à-vis established categories shapes audience attention, search, and evaluation. Studying how categorization shapes processes of attention and search in patent citations, for example, Kovács, Carnabuci, and Wezel (2021) find that an inventor's ability to identify an invention as relevant is facilitated by its classification within the boundaries of a distinctive and well-defined technological category. In many cases, easier identification coincides with greater appeal. For example, Zuckerman (1999) finds that firms that conform to categorization schemes held by financial analysts for sorting firms into reference groups are more likely to receive analyst coverage, leading to greater valuation in stock prices. In other cases, audience attention and appeal diverge. For instance, studying film projects, Hsu (2006) finds that films that span categories attract larger audiences due to their broader potential appeal, but also tend to achieve lower critical and audience ratings.

In this paper, we examine how the positioning of an author's books relative to prior offerings shapes the processes of search, attention, and appeal in distinct but interrelated ways. We find that, when authors attempt to reposition themselves through new products, there is likely to be a label mismatch—particularly when the audience for the new book consists of many repeat audience members who carry over prior identity expectations. This mismatch then leads to lowered attention and appeal.

Overall, categorical repositioning may result in two kinds of reception penalties. The first penalty concerns the amount of attention that a book attracts. When audiences label a new book differently than the position the author claimed, attention may suffer as readers quickly discount or discontinue consumption of the book if there is a mismatch between their ex ante expectations and what the book is delivering. Label mismatch also could mean that readers who might actually find the book's underlying content appealing will be less likely to recognize that the book fits with their own preferences due to its labeling by other readers. In other words, a greater label mismatch restricts the matching of market offerings with consumers whose preferences align with those offerings.

The second type of penalty is associated with repeat audience members, who are likely to pay attention to an author's new book because of their prior experience and knowledge of the author. Recent research has underscored how heterogeneity in audience characteristics shapes evaluative processes. For example, scholars have investigated the role of audiences' goals (Ertug et al. 2016, Paolella and Durand 2016, Pontikes 2012) and their taste for variety (Goldberg, Kovacs and Hannan 2016). Other studies have considered how ties between producers and professional evaluators shape evaluators' assessments with a focus on the presence of favoritism or conflicts of interest (Bowers and Prato 2018, Fleischer 2009, Olson and Waguespack 2020). Yet, we know little about how evaluator experience with a producer constrains the way in which new information is perceived and interpreted—an important issue for understanding the factors shaping the success of producers' repositioning efforts.

Prior experience with an author could make it more likely that a book will make it to a reader's consideration set even though its content may not actually fit with the reader's preferences. And the more a new book departs from prior books' content, the more likely repeat audience members are to label the book differently from the author's intended positioning. This dynamic may be exacerbated in experience-good settings such as the book industry, where consumers are unable to determine the qualities of a good prior to purchase (Nelson, 1970). Several readers we interviewed noted that they often follow authors whose work they enjoyed, and were at times surprised when subsequent books diverged substantially from what they had previously read. On websites where users' labels and ratings influence other readers, ending up in poorly matched consideration sets through label mismatch could ultimately harm a book's reputation and success. Future work is needed to better establish the relevance of these two types of reception penalties in market contexts.

More generally, our research complements studies that have pushed to move beyond an abstract and homogeneous view of the market audience towards thinking about the different preferences and motives of different audience types. For example, Pontikes (2012) draws a distinction between audiences like venture capitalists who are “market makers”—who try to redefine or make new categorizations—versus “market takers” like consumers who prefer typical offerings. Goldberg, Hannan, and Kovács (2016) distinguish between audience types based on the value they place on variety and typicality of market offerings with respect to established categories. Because of their different orientations and roles, these different audience types are expected to respond differently to offerings that span market categories at a single point in time. We extend this literature by considering how category constraints may play out over the sequence of careers. Our distinction between repeat versus new audience members highlights the differences in how audience members are influenced by market categories depending on the extent of their prior relationship with a given market producer.

Our findings also relate to work within literature on innovation, which has highlighted the critical role evaluation processes play in the identification of innovative, high-quality projects (Criscuolo et al. 2017, Lane et al. 2022). Studies suggest that the lenses that evaluators use to evaluate a project or proposal impact the way in which they judge its innovativeness or quality. For example, Boudreau et al. (2016) find evidence that highly novel proposals are often interpreted through existing knowledge frames, and as a result misconstrued and discounted. We find a distinct, although related process: when an audience member labels and interprets an author’s work through a category frame that is different from the author’s intended positioning, they are likely to feel disappointment and discount the work. One implication is that evaluators who adopt a different knowledge frame than what innovators intend will be less likely to appreciate a proposal’s potential impact and quality, a process that has significant implications for the progression of knowledge when the evaluators of interest are key gatekeepers to funding and project sponsorship.

Another general body of literature that our study speaks to concerns change in organizational identities. Deviating from an existing organizational identity can result in disapproval from audiences who have formed expectations and may view these deviations as illegitimate or a violation of prior commitments (Hsu and Hannan 2005). Organizational identity researchers have studied micro-level factors that contribute to resistance against organizational change efforts. Changes in organizational identity creates anxiety, conflict, and loss of self-esteem among individuals such as employees or even highly devoted consumers whose own identity is linked to the organization's identity in some way (Ashforth and Mael 1989, Gioia et al. 2010a). When organizations attempt major changes, actions they take that are inconsistent with existing organizational identities are difficult for members to understand and challenge their own identities (Reger et al. 1994).

In addition to these more motivated reasons that audiences resist change, we find that existing identity expectations systematically bias the way information is perceived, reducing the ability of audiences to process information inconsistent with prior identities held and leading to label mismatch. This is consistent with Tripsas' (2009) and Benner's (2010) observations of cognitive inertia in analysts' perceptions of organizational change efforts. Through our empirical findings, we document the processes through which this inertia impacts different stages of audience reception. We also extend the understanding of this phenomenon by showing how particular kinds of exchange relationships and audience preference profiles shape the way categorical stickiness influences market processes and outcomes.

Methodological contribution. Organizational scholars have long been interested in understanding the extent to which the positioning of producers in one or more categories relates to their valuation by social actors (e.g., Zuckerman, 1999; Hsu, 2006; Hsu, Hannan, and Koçak, 2009). Relatedly, positioning, categorization, and labeling have been the focus of a sizeable research stream, investigating topics and settings such as wine (Negro and Leung 2013), restaurants (Kovács and Hannan 2015), careers (Ferguson and Hasan 2013, Leahey et al. 2017, Leung 2014, Leung and Sharkey 2014), and innovation and patents (Kovács et al. 2021). While insightful, this stream of research suffers from a common methodological limitation: researchers typically use observed category labels assigned to producers and/or offerings to assess their positioning, and they generally do not disentangle the endogenous effects of a product's features on how it is labeled (see Pontikes and Hannan, 2014; Kovács and Johnson, 2014 for exceptions and see Negro and Leung 2013 and Leung and Sharkey 2014 for ways of addressing this issue though research designs involving natural experiments).

This is problematic for two interrelated reasons. First, taking observed labels as a proxy for underlying content and positioning makes it difficult to separate “content” effects from “labeling” effects. For example, it is not possible to disentangle whether a category-spanning restaurant receives lower ratings because it serves a diverse set of dishes or because it is labelled as category-spanning. Second, the process of assigning labels to organizations is one that is socially and cognitively biased. For example, the observed labels of books are likely to be a better proxy of books by newer authors than those by experienced ones. This finding highlights that it would be important for researchers to measure producers’ intended positioning in content space separately from positioning in label space. While previous methods and data types did not allow for separate treatment of each space, we believe that text data and NLP, deep learning, and other AI methods present great potential to move the field forward. In this paper, we demonstrated how deep learning can aid in locating books in a multidimensional space based on their text, independently of their assigned labels. This approach could be extended to other settings, data sources, and research questions. For example, deep learning could be used to analyze content within press release data, Twitter feeds, restaurant menus, or quarterly earnings. It could also be used to examine effects of category positioning more precisely, labeling, and optimal distinctiveness in a variety of contexts.

Limitations.

There are limitations to our study. First, we do not study category spanning, but rather look at the averaged genre profile positions of books in our dataset. We study moves over time in this averaged profile and how they impact the reception of an author’s book. Our focus in the current paper is on moves within genre space over time rather than category spanning at a given point in time. Future research is needed to understand how category spanning at a given point in time interacts with temporal positioning.

Second, our study uses ML-based measures to capture key constructs, such as repositioning and label mismatch, in a nuanced manner. We aimed to bolster the accuracy of the prediction model used to generate these measures by i) training the model using data on the synopses and labels of books by authors who only wrote one book, thereby eliminating the possibility that authors' prior work would have affected users' labeling behavior and ii) employing a moving window approach to training, thus ensuring that our measures correctly reflect any changes in the meaning of genres over time. We have also sought to validate these measures in multiple ways, such as (i) examining out-of-sample prediction performance, and (ii) by collecting human classification in a controlled environment and showing that the model predictions are not biased by the author's level of repositioning. Nonetheless, it is important to note that ML-based measures are prone to various limitations and challenges, such as the possibility that other unmeasured facets of the text could confound the measured treatment's effect and the potential that the training data does not fully encapsulate the entity that we wish to measure (i.e., we train on synopses of books rather than the full text of the books) (Fong and Grimmer 2023). These limitations suggest that a degree of caution is warranted in interpreting our results.

In addition, it is important to acknowledge the correlational nature of our findings. To bolster our ability to draw causal inferences, we conducted tests in which we, for example, held constant the underlying content of a book by examining readers who read the same book after reading a different first book. To some extent, tests of that nature also help mitigate concerns about the endogeneity of an author's decision to reposition. Ultimately, however, the processes by which authors choose to reposition, and readers choose to read books in a certain order or to label them a certain way, are opaque to us, limiting our ability to draw causal conclusions. Despite this, we believe there is substantial value in studying the process of mislabeling and its impacts using observational data with high external validity. We hope that our work stimulates further exploration and examination using different research designs and data sources.

References

Anon About Goodreads. Retrieved (December 24, 2020), https://www.goodreads.com/about/us.

Ashforth BE, Mael F (1989) Social identity theory and the organization. *Acad. Manage. Rev.* 14:20–39.

Barnett WP, Pontikes EG (2008) The Red Queen, success bias, and organizational inertia. *Manag. Sci.* 54(7):1237–1251.

Becker HS (1982) *Art worlds.* University of California Press, Berkeley; Los Angeles; London.

Benner MJ (2010) Securities analysts and incumbent response to radical technological change: Evidence from digital photography and internet telephony. *Organ. Sci.* 21(1):42–62.

Blackwell M, Iacus S, King G, Porro G (2009) Cem: Coarsened exact matching in Stata. *Stata J.* 9(4):524–546.

Boudreau KJ, Guinan EC, Lakhani KR, Riedl C (2016) Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Manag. Sci.* 62(10):2765–2783.

Bowers A, Prato M (2018) The structural origins of unearned status: How arbitrary changes in categories affect status position and market impact. *Adm. Sci. Q.* 63(3):668–699.

Bransford N (2018) Why it's important to know your book's genre. *Nathan Bransford Writ. Book Ed. Publ.* Retrieved (December 24, 2020), https://nathanbransford.com/blog/2018/04/why-its-important-to-know-your-genre.

Carnabuci G, Operti E, Kovács B (2015) The categorical imperative and structural reproduction: Dynamics of technological entry in the semiconductor industry. *Organ. Sci.* 26(6):1734–1751.

Chatterji AK (2009) Spawned with a silver spoon? Entrepreneurial performance and innovation in the medical device industry. *Strateg. Manag. J.* 30(2):185–206.

Childress C (2017) *Under the Cover: The Creation, Production, and Reception of a Novel* (Princeton University Press).

Chollet F (2017) *Deep Learning with Python* 1st edition. Manning Publications, Shelter Island, New York.

Criscuolo P, Dahlander L, Grohsjean T, Salter A (2017) Evaluating novelty: The role of panels in the selection of R&D projects. *Acad. Manage. J.* 60(2):433–460.

Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. *ArXiv181004805 Cs*.

Dobrev SD (2007) Competing in the looking-glass market: Imitation, resources, and crowding. *Strateg. Manag. J.* 28(13):1267–1289.

Dobrev SD, Kim T, Hannan MT (2001) Dynamics of niche width and resource partitioning. *Am. J. Sociol.* 106(5):1299–1337.

Durand R, Rao H, Monin P (2007) Code and conduct in French cuisine: Impact of code changes on external evaluations. *Strateg. Manag. J.* 28(5):455–472.

Eggers JP, Song L (2015) Dealing with failure: Serial entrepreneurs and the costs of changing industries between ventures. *Acad. Manage. J.* 58(6):1785–1803.

Ertug G, Yogev T, Lee YG, Hedström P (2016) The art of representation: How audience-specific reputations affect success in the contemporary art field. *Acad. Manage. J.* 59(1):113–134.

Faulkner RR (1983) *Music on Demand*. Transaction Publishers.

Ferguson JP, Hasan S (2013) Specialization and career dynamics: Evidence from the Indian Administrative Service. 58(2):233–256.

Fleischer A (2009) Ambiguity and the equity of rating systems: United States brokerage firms, 1995–2000. *Adm. Sci. Q.* 54(4):555–574.

Fong C, Grimmer J (2023) Causal inference with latent treatments. *Am. J. Polit. Sci.* 67(2):374–389.

Gioia DA, Price KN, Hamilton AL, Thomas JB (2010a) Forging an identity: An insider-outsider study of processes involved in the formation of organizational identity. *Adm. Sci. Q.* 55(1):1–46.

Glynn MA, Abzug R (2002) Institutionalizing identity: Symbolic isomorphism and organizational names. *Acad. Manage. J.* 45(1):267–280.

Glynn MA, Navis C (2013) Categories, identities, and cultural classification: Moving beyond a model of categorical constraint: Categories, identities, and cultural classification. *J. Manag. Stud.* 50(6):1124–1137.
Godart F, Hsu G, Negro G (2023) Gatekeeping and the use of contested practices in creative industries: The case of fur in fashion. *Organ. Sci.*34(2):637-656.
Gold J (2014) The pros and cons of switching genres — Guest: Summerita Rhayne. *Jami Gold Paranorm. Author,* Nov. 6. http://www.jamigold.com/2014/11/the-pros-and-cons-of-switching-genres-guest-summerita-rhayne/.
Goldberg A, Hannan MT, Kovács B (2016) What does it mean to span cultural boundaries? Variety and atypicality in cultural consumption. *Am. Sociol. Rev.*81(2):215-241.
Goodfellow I, Bengio Y, Courville A (2016) *Deep Learning*, Illustrated edition. The MIT Press, Cambridge, Massachusetts.
Granqvist N, Grodal S, Woolley JL (2013) Hedging your bets: Explaining executives' market labeling strategies in nanotechnology. *Organ. Sci.* 24(2):395–413.
Hannan MT, Carroll GR, Pólos L (2003) The organizational niche. *Sociol. Theory* 21(4):309–340.
Haveman HA, Nonnemaker L (2000) Competition in multiple geographic markets: The impact on growth and market entry. *Adm. Sci. Q.* 45(2):232.
Hayes AF (2009) Beyond Baron and Kenny: Statistical mediation analysis in the new Millennium. *Commun. Monogr.* 76(4):408–420.
Hirsch PM (1972) Processing fads and fashions: An organization-set analysis of cultural industry systems. *Am. J. Sociol.* 77(4):639–659.
Hsu G (2006) Jacks of all trades and masters of none: Audiences' reactions to spanning genres in feature film production. *Adm. Sci. Q.* 51(3):420–450.
Hsu G, Hannan MT (2005) Identities, genres, and organizational forms. *Organ. Sci.* 16(5):474–490.
Hsu G, Hannan MT, Koçak Ö (2009) Multiple category memberships in markets: An integrative theory and two empirical tests. *Am. Sociol. Rev.* 74(1):150–169.
Ibarra H (2003) *Working Identity: Unconventional strategies for reinventing your career*. Harvard Business School Press, Cambridge, MA.
Ibarra H, Barbulescu R (2010) Identity as narrative: Prevalence, effectiveness, and consequences of narrative identity work in macro work role transitions. *Acad. Manage. Rev.* 35(1):135–154.
Jones C, Svejenova S, Pedersen JS, Townley B (2016) Misfits, mavericks and mainstreams: Drivers of innovation in the creative industries. *Organ. Stud.* 37(6):751–768.
Kaplan S, Tripsas M (2008) Thinking about technology: Applying a cognitive lens to technical change. *Res. Policy* 37(5):790–805.
Kennedy MT (2008) Getting counted: Markets, media, and reality. *Am. Sociol. Rev.* 73(2):270–295.
Kovács B, Carnabuci G, Wezel FC (2021) Categories, attention, and the impact inventions. *Strateg. Manag. J.* 42(5):992-1023.
Kovács B, Hannan MT (2015) Conceptual spaces and the consequences of category spanning. *Sociol. Sci.* 2:252–286.
Kovács B, Johnson R (2014) Contrasting alternative explanations for the consequences of category spanning: A study of restaurant reviews and menus in San Francisco. *Strateg. Organ.* 12(1):7–37.
Lahiri A, Wadhwa A (2021) When do serial entrepreneurs found innovative ventures? Evidence from patent data. *Small Bus. Econ.*57(4):1973-1993.
Lane JN, Teplitskiy M, Gray G, Ranu H, Menietti M, Guinan EC, Lakhani KR (2022) Conservatism gets funded? A field experiment on the role of negative information in novel project evaluation. *Manag. Sci.* 68(6):4478–4495.
Le Mens G, Kovács B, Hannan MT, Pros G (2023) Using machine learning to uncover the semantics of concepts: How well do typicality measures extracted from a BERT text classifier match human judgments of genre typicality? *Sociol. Sci.* 10:82–117.
Leahey E, Beckman CM, Stanko TL (2017) Prominent but less productive: The impact of interdisciplinarity on scientists' research*. *Adm. Sci. Q.* 62(1):105–139.

Leung MD (2014) Dilettante or renaissance person? How the order of job experiences affects hiring in an external labor market. *Am. Sociol. Rev.* 79(1):136–158.
Leung MD, Sharkey AJ (2014) Out of sight, out of mind? Evidence of perceptual factors in the multiple-category discount. *Organ. Sci.* 25(1):171–184.
Lounsbury M, Glynn MA (2001) Cultural entrepreneurship: stories, legitimacy, and the acquisition of resources. *Strateg. Manag. J.* 22(6–7):545–564.
Nayak P (2019) Understanding searches better than ever before. *Google*. Retrieved (February 26, 2021), https://blog.google/products/search/search-language-understanding-bert/.
Negro G, Kovács B, Carroll GR (2022) What's next? Artists' music after Grammy awards. *Am. Sociol. Rev.* 87(4):644–674.
Negro G, Leung MD (2013) "Actual" and perceptual effects of category spanning. *Organ. Sci.* 24(3):684–696.
Olson DM, Waguespack DM (2020) Strategic behavior by market intermediaries. *Strateg. Manag. J.* 41(13):2474–2492.
O'Mahony S, Bechky BA (2006) Stretchwork: Managing the career progression paradox in external labor markets. *Acad. Manage. J.* 49(5):918–941.
Paolella L, Durand R (2016) Category spanning, evaluation, and performance: Revised theory and test on the corporate law market. *Acad. Manage. J.* 59(1):330–351.
Peterson RA, Kern RM (1996) Changing highbrow taste: From snob to omnivore. *Am. Sociol. Rev.* 61(5):900–907.
Pontikes E, Hannan M (2014) An ecology of social categories. *Sociol. Sci.* 1:311–343.
Pontikes EG (2012) Two sides of the same coin: How ambiguous classification affects multiple audiences' evaluations. *Adm. Sci. Q.* 57(1):81–118.
Porter ME (1980) *Competitive Strategy: Techniques for analyzing industries and competitors*, Free Press.
Ravasi D, Phillips N (2011) Strategies of alignment: Organizational identity management and strategic change at Bang & Olufsen. *Strateg. Organ.* 9(2):103–135.
Reger RK, Gustafson LT, Demarie SM, Mullane JV (1994) Reframing the organization: Why implementing total quality is easier said than done. *Acad. Manage. Rev.* 19(3):565–584.
Rindova VP, Williamson IO, Petkova AP, Sever JM (2005) Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation. *Acad. Manage. J.* 48(6):1033–1049.
Ruef M, Patterson K (2009) Credit and classification: The impact of industry boundaries in Nineteenth-century America. *Adm. Sci. Q.* 54(3):486–520.
Schimmer M, Brauer M (2012) Firm performance and aspiration levels as determinants of a firm's strategic repositioning within strategic group structures. *Strateg. Organ.* 10(4):406–435.
Sharkey A, Kovács B, Hsu G (2023) Expert critics, rankings, and review aggregators: The changing nature of intermediation and the rise of markets with multiple intermediaries. *Acad. Manag. Ann.* 17(1):1–36.
Smith EB (2011) Identities as lenses: How organizational identity affects audiences' evaluation of organizational performance. *Adm. Sci. Q.* 56(1):61–94.
Tripsas M (2009) Technology, identity, and inertia through the lens of "The Digital Photography Company." *Organ. Sci.* 20(2):441–460.
Washington M, Zajac EJ (2005) Status evolution and competition: Theory and evidence. *Acad. Manage. J.* 48(2):282–296.
Whetten DA (2006) Albert and Whetten revisited: Strengthening the concept of organizational identity. *J. Manag. Inq.* 15(3):219–234.
White HC (1981) Where do markets come from? *Am. J. Sociol.* 87(3):517–547.
Younkin P, Kashkooli K (2020) Stay true to your roots? Category distance, hierarchy, and the performance of new entrants in the music industry. *Organ. Sci.* 31(3):604–627.
Zuckerman EW (1999) The categorical imperative: securities analysts and the illegitimacy discount. *Am. J. Sociol.* 104(5):1398–1438.

Zuckerman EW, Kim T, Ukanwa K, von Rittmann J (2003) Robust identities or nonentities? Typecasting in the feature-film labor market. *Am. J. Sociol.* 108(5):1018–1073.

Tables

Table 1. Illustration of the observed and the predicted label proportions for the book "Derailed" by James Siegel. Link: https://www.goodreads.com/book/show/314393

Book description: "Advertising director Charles Schine is just another New York commuter, regularly catching the 8.43 to work. But the day he misses his train is the day that changes his life. Catching the 9.05 instead, he can't help but be drawn by the sight of the person opposite. Charles has never cheated on his wife in eighteen years of marriage. But then Charles has never met anyone like Lucinda Harris before. Charming, beautiful and a seductively good listener, Charles finds himself instantly attracted. And though Lucinda is married too, it is immediately apparent that the feeling is mutual. Their journeys into work become lunch dates, which become cocktails and eventually lead to a rented room in a seedy hotel. They both know the risks they are taking, but not in their worst nightmares could they foresee what is to follow. Suddenly their temptation turns horrifically sour, and their illicit liaison becomes caught up in something bigger, more dangerous, more brutally violent. Unable to talk to his partner or the police, Charles finds himself trapped in a world of dark conspiracy and psychological games. Somehow he's got to find a way to fight back, or his entire life will be spectacularly derailed for good."

Observed and predicted label proportions

Label	Observed	Predicted	Label	Observed	Predicted
biography	0.000	0.001	memoir	0.000	0.001
business	0.000	0.000	music	0.000	0.001
chicklit	0.000	0.004	mystery	0.129	0.193
childrens	0.000	0.000	nonfiction	0.000	0.004
christian	0.000	0.003	paranormal	0.000	0.002
classics	0.000	0.001	philosophy	0.000	0.000
comics	0.000	0.001	poetry	0.000	0.000
contemporary	0.000	0.064	psychology	0.000	0.002
cookbooks	0.000	0.000	religion	0.000	0.001
crime	0.100	0.059	romance	0.000	0.102
fantasy	0.000	0.002	science	0.000	0.000
fiction	0.286	0.313	sciencefiction	0.000	0.002
gayandlesbian	0.000	0.007	selfhelp	0.000	0.000
graphicnovels	0.000	0.002	spirituality	0.000	0.000
historicalfiction	0.000	0.009	sports	0.000	0.000
history	0.000	0.005	suspense	0.143	0.077
horror	0.000	0.004	thriller	0.343	0.126
humorandcomedy	0.000	0.007	travel	0.000	0.002
manga	0.000	0.000	youngadult	0.000	0.002

Table 2. Overview of the hypotheses and tests

Hypotheses		Main results (book level)	Alternative specification (individual level)	Post hoc supplementary findings
H1	The more a producer repositions from prior offerings to the current offering, the greater the mismatch between the producer's positioning and the audience's labels of its current offering.	Table 4, book level model	Appendix Table A6, individual level mismatch	Figure 3, plotting label mismatch as a function of author repositioning; Appendix Table A7, weaker effect with higher author diversity
H2	Label mismatch will be greater for audience members who have greater prior exchange or consumption experience with the producer.	Table 4, book level model		
H3	The greater the label mismatch, the less attention audience members will give it and the less appealing they will find it to be.	Table 5, matched book level sample	Table 6: individual level label mismatch and appeal; Appendix Table A8, matched book level sample with continuous measure of label mismatch; Appendix Table A9, matched book level sample with label mismatch split at terciles	Appendix Table A10: matched book level sample, split to new vs. repeat readers; Appendix Table A11: reader omnivorousness models; Appendix Table A12: negative binomial count models

Table 3. Descriptive statistics and pairwise correlations

3a: Sample: Second and later books of authors, for results presented in Table 4

Variable	Mean	SD	Min	Max	(1)	(2)	(3)
1. Label mismatch	0.376	0.196	0.003	1.324			
2. Extent of repositioning	0.188	0.214	0	1.831	0.115		
3. Repeat readership	0.468	0.302	0.0001	1	0.033	-0.015	
4. Author book order (ln)	2.111	1.080	0.693	6.418	0.038	0.021	0.170

N= 308,795 2nd and later books by authors with at least 2 books

3b: Matched sample, for results presented in Table 5

Variable	Mean	Std. dev.	Min	Max	(1)	(2)	(3)	(4)	(5)
1. Rating count of book (ln)	3.433	1.472	0.693	12.635					
2. 5-star rating count of book (ln)	2.228	1.475	0.000	12.225	0.935				
3. Mean rating	3.820	0.451	1.000	5.000	0.049	0.324			
4. Label mismatch high (above median)	0.639	0.480	0.000	1.000	-0.068	-0.060	-0.012		
5. Repeat readership	0.468	0.357	0.000	1.000	-0.389	-0.333	0.026	0.009	
6. Extent of repositioning	0.146	0.198	0.000	1.801	-0.048	-0.050	-0.011	0.055	-0.020

N=262,503 books

Table 4. Book-level estimates of label mismatch, for authors' second and later books

DV: label mismatch	(1)	(2)	(3)
Extent of repositioning	0.0877***	0.0877***	0.0749***
	-0.00172	-0.00172	-0.00288
Repeat readership		0.00182	-0.00426*
		-0.00146	-0.00182
Repeat readership X			0.0279***
Extent of repositioning			-0.00503
Author book order (ln)	-0.0122***	-0.0123***	-0.0121***
	-0.000759	-0.000764	-0.000765
Total labels applied to focal book (ln)	-0.0308***	-0.0307***	-0.0307***
	-0.000339	-0.000362	-0.000362
Publication year of focal book	-0.00351***	-0.00351***	-0.00351***
	-0.0000781	-0.0000781	-0.0000781
Mean rating of author's prior books	0.00118	0.00108	0.00114
	-0.00347	-0.00347	-0.00347
Constant	7.537***	7.535***	7.544***
	-0.157	-0.157	-0.157
Author FE	Y	Y	Y
N	308795	308795	308795
adj. R-sq	0.246	0.246	0.246
Log likelihood	130050.3	130051.2	130069

Standard errors (clustered on authors) in parentheses.
* p<0.05, **p<0.01, ***p<0.001
Notes: The sample is constructed by starting with the second and later books from all authors who have written two or more books. Our use of author fixed effects requires that we have at least two books from an author.

Table 5. Book-level estimates of total Goodreads ratings counts and mean ratings assigned, matched sample+

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	DV: Total count of ratings (ln)			DV: Count of 5-star ratings (ln)			DV: Mean rating		
Label mismatch (high)	-0.187***		-0.175***	-0.199***		-0.187***	-0.0362***		-0.0360***
	-0.0143		-0.0135	-0.0147		-0.0142	-0.00472		-0.00493
Repeat readership	-1.603***	-1.596***	-1.604***	-1.371***	-1.363***	-1.372***	0.0660***	0.0689***	0.0659***
	-0.0195	-0.0198	-0.0195	-0.0225	-0.0229	-0.0226	-0.00759	-0.00742	-0.00748
Extent of repositioning		-0.433***	-0.403***		-0.440***	-0.407***		-0.0134	-0.0075
		-0.0309	-0.0336		-0.0282	-0.0298		-0.015	-0.0147
Total count of ratings (ln)							0.0225***	0.0232***	0.0224***
							-0.00163	-0.00163	-0.00161
Constant	4.302***	4.242***	4.354***	2.997***	2.930***	3.049***	3.735***	3.710***	3.736***
	-0.0164	-0.017	-0.0185	-0.0175	-0.0172	-0.0188	-0.00803	-0.00879	-0.00796
N	262503	262503	262503	262503	262503	262503	262503	262503	262503
adj. R-sq	0.153	0.153	0.156	0.113	0.112	0.116	0.007	0.006	0.007
Log likelihood	-452027.8	-452078.4	-451572.4	-458835.2	-458940.4	-458393.4	-162431.9	-162622.6	-162430.5

Standard errors (clustered on matched strata) in parentheses
* p<0.05, **p<0.01, ***p<0.001
Notes: Treatment is based on above- versus below-median values of label mismatch. The sample is constructed by starting with all books from all authors who have written two or more books and creating a matched sample that pairs each book with a high value of label mismatch with a book that has a low value of label mismatch but is observationally equivalent in terms of the following metrics: deciles of the total count of ratings of all prior books by the author, deciles of mean rating of all prior books by the author, deciles of the focal book's publication year, exact count of prior books published by the author, and the most common predicted genre label of the focal book.

Table 6. Individual-level estimates of Goodreads ratings assigned.

	Fixed Effects Specification				Matched Sample		
DV: individual rating	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Individual-level label mismatch	-0.0520***	-0.0490***	-0.0519***	-0.0373***	-0.0499*	-0.0487*	-0.035
	-0.00515	-0.00582	-0.00514	-0.00423	-0.0203	-0.0203	-0.0188
Repeat reader		0.00925	0.00292				
		-0.00612	-0.00377				
Repeat reader X Label mismatch		-0.00635					
		-0.0065					
Individual-level repositioning			0.599***			-0.113**	
			-0.105			-0.0417	
Repeat reader X Individual-level repositioning			-0.579***				
			-0.105				
Individual-level predicted rating				0.814***			0.818***
				-0.00319			-0.0103
Review year	-0.00746*	-0.00751*	-0.00752*	-0.00738*	-0.00143	-0.00133	-0.00321
	-0.00291	-0.00296	-0.00296	-0.00288	-0.00285	-0.00285	-0.00262
Constant	18.93**	19.02**	19.03**	15.65**	6.591	6.416	7.14
	-5.853	-5.948	-5.953	-5.792	-5.737	-5.737	-5.269
User FE	Y	Y	Y	Y	Y	Y	Y
Book FE	Y	Y	Y	Y	Y	Y	Y
N	1285992	1285992	1285992	1284332	64627	64627	64627
adj. R-sq	0.314	0.314	0.314	0.392	0.301	0.301	0.391
Log lik.	-1404773.3	-1404769.4	-1404745.5	-1325394.1	-62339.1	-62333.7	-57898.5

Standard errors (clustered on users and books) in parentheses

* p<0.05, **p<0.01, ***p<0.001

Notes: Sample for FE models (Models 1-4) is constructed by examining user-level ratings and labels for all books from all authors who have written two or more books. (A user must have both rated the book and labeled it to enter the sample.) To create the matched sample (Models 5-7), we further culled the data to focus on pairs of users who had read the same second book by an author but a different first book. We analyze ratings of these second books as a function of variation in repositioning from the first book and label mismatch of the second book.

FIGURES

Figure 1. Out-of-sample prediction accuracy (averaged across all 38 labels). Based on N= 15,339,688 label-book pair observations (403,676 books * 38 labels). Note that there is no observation at x=1 because there is no book for which any of the labels by itself would be predicted to be more than 0.975.

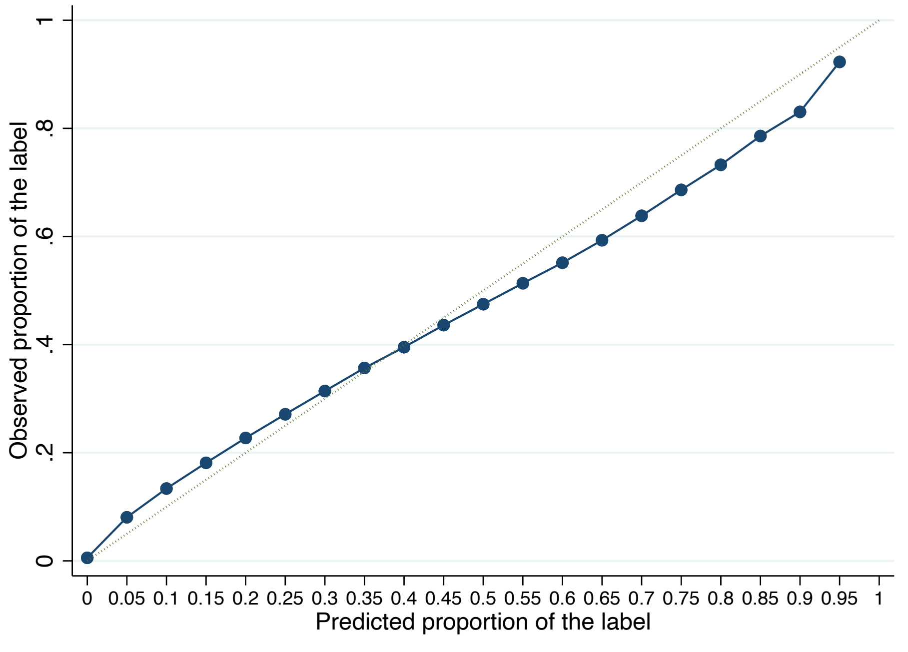


Figure 2. Example of individual-level matching approach

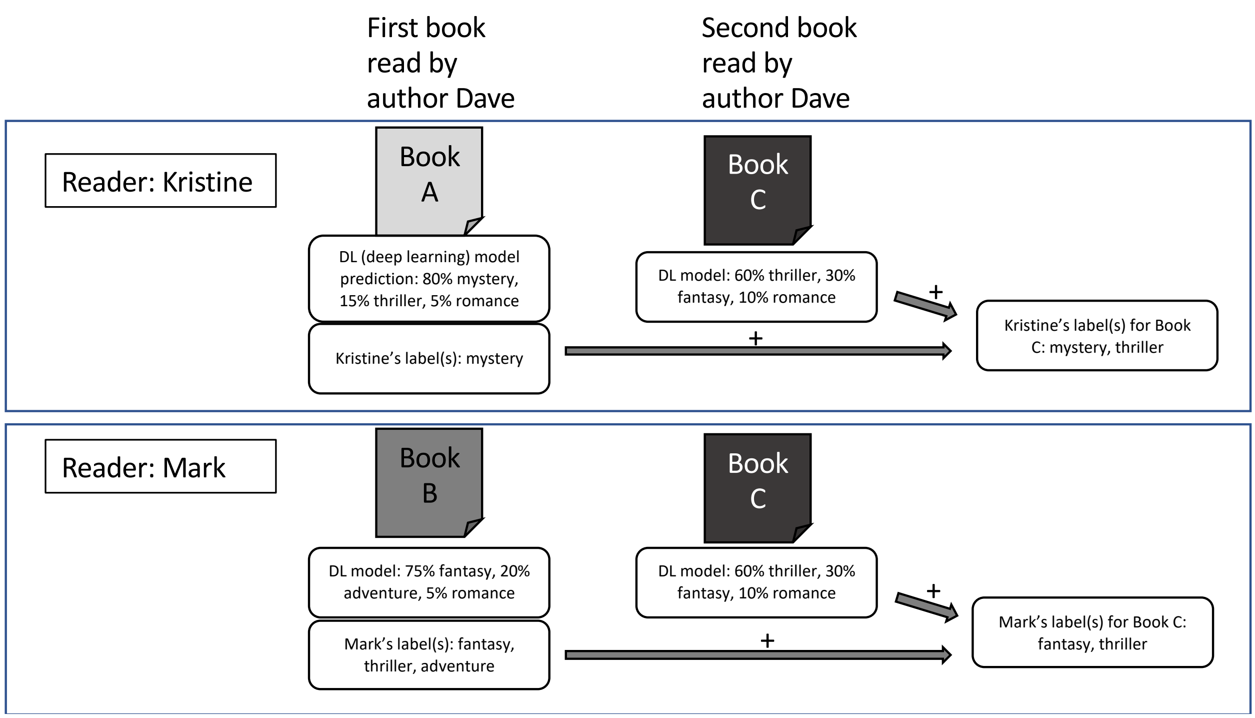
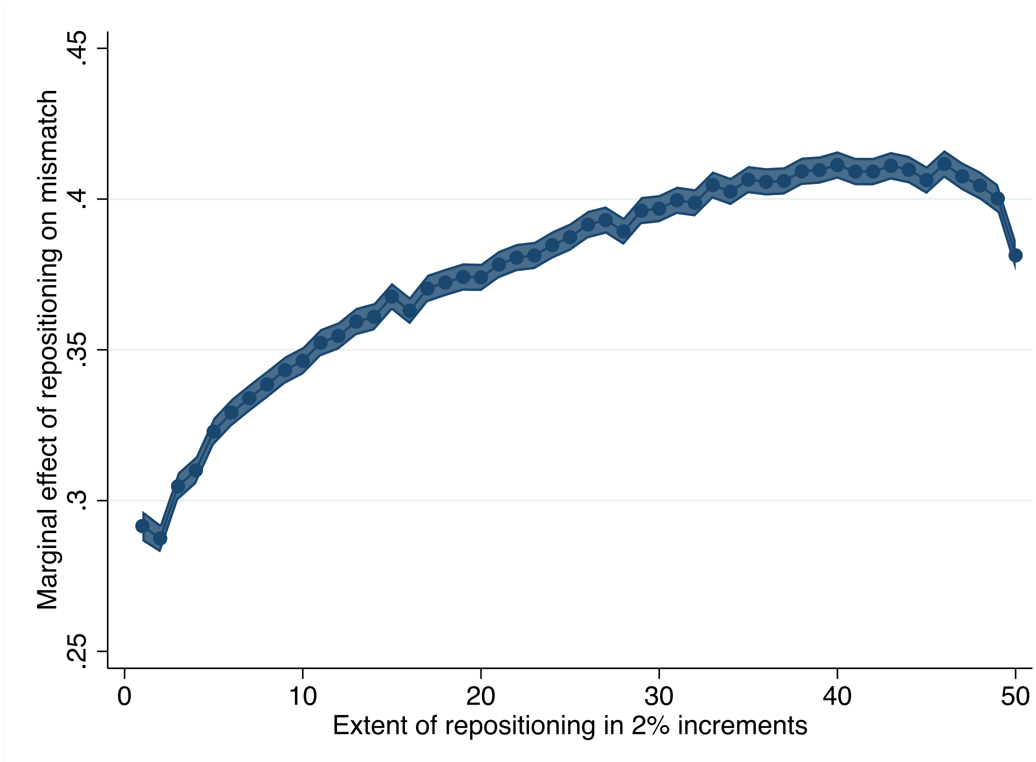


Figure 3. Average label mismatch, as a function of the book's claimed content-based distance from previous book(s) of the author (extent of repositioning). Marginal effects graph with author fixed effects.



Appendix A

Appendix Tables

Appendix Figures

Table A1 Genre classifications on Goodreads.com. Distribution of the observed labels for the books in the out-of-sample analyses

Label		Label	Observed
biography	1.17%	memoir	0.83%
business	0.31%	music	0.23%
chicklit	1.15%	mystery	4.96%
childrens	4.61%	nonfiction	6.24%
christian	0.42%	paranormal	2.91%
classics	0.53%	philosophy	0.42%
comics	2.63%	poetry	0.78%
contemporary	2.94%	psychology	0.38%
cookbooks	0.50%	religion	0.47%
crime	0.88%	romance	8.04%
fantasy	12.12%	science	0.75%
fiction	10.60%	sciencefiction	5.48%
gayandlesbian	0.33%	selfhelp	0.25%
graphicnovels	3.06%	spirituality	0.16%
historicalfiction	3.43%	sports	0.28%
history	3.59%	suspense	0.72%
horror	1.59%	thriller	1.23%
humorandcomedy	1.86%	travel	0.26%
manga	2.98%	youngadult	10.90%

N=403,676 (403676=65,555 first books and 338,121 second and greater books)

Table A2 Descriptives for supplementary (individual level) models

For the full sample where user rated at least two books from the authors

Variable	Mean	Std. dev.	Min	Max	1	2	3	4	5	6
1. Rating	3.864	0.944	1.000	5.000						
2. Individual-level label mismatch	0.746	0.265	0.003	1.410	-0.014					
3. Count of author's books rated by individual	2.239	3.920	1	208	0.013	-0.003				
4. Review year	2013.997	3.625	2006	4732	0.036	-0.017	0.036			
5. Individual-level extent of repositioning	0.078	0.152	0.000	1.861	-0.007	0.043	0.188	0.031		
6. Repeat reader	0.476	0.499	0	1	-0.011	0.004	0.331	0.063	0.538	
7. Predicted rating	3.786	0.533	-0.783	6.854	0.539	-0.030	0.018	0.050	0.009	0.005

N= 1,285,992

For the matched sample where user rated the same second book from the focal author

Variable	Mean	Std. dev.	Min	Max	1	2	3	
1. Rating	3.683	0.956	1.000	5.000				
2. Individual-level label mismatch	0.731	0.253	0.005	1.408	-0.030			
3. Review year	2014	2	2006	2017	0.055	-0.031		
4. Individual-level extent of repositioning	0.127	0.141	0.000	1.776	0.006	0.0563	0.037	
5. Predicted rating	3.722	0.530	0.400	5.938	0.546	-0.033	0.049	0.018

N= 64,627

Table A3 Balance table for matched book-level sample for H3

		Low_reposition	High_reposition	T-value	Cohen's d
Focal book's publication year	Mean	2007.038	2005.064	36.37	.142
	SD	9.260	10.416		
Prior book(s)' mean rating	Mean	3.853	3.836	17.18	.067
	SD	.248	.259		
Prior book(s)' rating count (ln)	Mean	5.194	5.219	-1.75	.007
	SD	3.652	3.618		

Note: We additionally required an exact match on the highest probability label for the author's current book and the count of prior books published by the author.

Table A4. Book/label level estimates of the category labels assigned by Goodreads users to a book, for authors' second and later books

DV: Proportion of labels observed	(1)	(2)	(3)
Label predictions based on focal book's synopsis	0.483***	0.480***	0.480***
	(0.0295)	(0.0295)	(0.0295)
Labels applied to an author's prior books	0.599***	0.395***	0.499***
	(0.03)	(0.0366)	(0.033)
Total labels applied to focal book, ln	0.0000113	0.0000169	0.0000142
	(0.000299)	(0.000303)	(0.000309)
Publication year of focal book	-0.0000538	-0.0000607	-0.0000542
	(0.0000308)	(0.0000321)	(0.0000351)
Cumulative rating count of author's prior books (ln)		- 0.000663***	
		(0.00017)	
Labels applied to an author's prior books X		0.0274***	
Cumulative rating count of author's prior books (ln)		(0.00229)	
Author book order (ln)			-0.00140**
			(0.000423)
Labels applied to an author's prior books X			0.0549***
Author book order (ln)			(0.00708)
Constant	0.106	0.125	0.11
	(0.0624)	(0.0652)	(0.0709)
Author FE	Y	Y	Y
Label FE	Y	Y	Y
Adj. R-sq	0.74	0.744	0.743
Log likelihood	19250166.8	19343482.3	19321604.6

N=12,848,598

Standard errors (clustered on authors and labels) in parentheses; * p<0.05, **p<0.01, ***p<0.001

Notes: Label predictions based on a producer's synopsis. The variable is the output of the deep learning model that predicts how the book would be labelled based on its synopsis alone. It is calculated for each book-label pairing and it contains the proportion of the focal label (as compared to all labels) the book is predicted to be assigned based on its synopsis. For example, for a given book, the model may predict 70% thriller, 30% adventure and 0% for all other labels.

Labels applied to an author's previous books. This variable is used to capture categorical stickiness. It is calculated for each book-label pair and reflects the number of times the focal label was assigned to the book out of all the times it was labelled, averaged across all of the author's previous books. For example, for the third book of the author and the label fantasy, we calculate the observed proportion of fantasy labels for the author's first and second books and average these two values.

Table A5 Individual/label level estimates of category labels assigned to a book

One potential concern with the book-level test is that our deep learning algorithm may not fully account for unobserved factors shaping genre label assignments, which could lead to a spurious correlation between observed labels for an author's prior and current books. To address this, we conduct results at the individual reader level, using both a) the full sample of all readers for whom we have individual-level labelling data, as well as b) the sample of paired readers of authors' second books. These results compare labels assigned by individuals who have read the same second book by an author after each reading a different book by that same author.

For example, consider the following hypothetical case (depicted in Figure 2): there are two reviewers (Kristine and Mark), who have each read a different book by author Dave. Kristine read Book A, which she labeled as "mystery", while Mark read Book B, which he labeled as "fantasy", "thriller", and "adventure". Both read the same second book (Book C) by Dave, classified by our deep learning model based on the book's synopsis as 60% thriller, 30% fantasy, and 10% romance. We expect that Kristine's labeling of Book C will be influenced both by the label she applied to Book A and the deep learning model's prediction for Book C—the result is a label of "mystery" and "thriller". We predict a similar dynamic for Mark. Because the labels Mark assigned to Book A are closer to the synopsis-based prediction of Book than the labels Kristine assigned to Book B, Mark's labeling of Book C is expected to be closer to the model's prediction than Kristine's. That is, Mark's individual-level label mismatch for Book C will be smaller than Kristine's.

The results in Table A5 are consistent with this prediction. The label an individual reader assigns to the second book read by an author is significantly influenced by both the label assigned by the reader to the first book they read by that author and the label predictions based on the second book's synopsis. Given that the second book is the same (and we estimate book FEs), the fact that a significant remainder of the variance in labels can be explained by the labels the reader gave to the first book the reader consumed suggests the results are not simply due to prediction error of the deep learning algorithm.

	(1)	(2)
DV: Proportion of labels observed, individual level	Sample: All books (reader order>=2) for which there are individual reader labels	Sample: Second reviews in the matched reader sample
Label predictions based on focal book's synopsis	0.194***	0.283***
	(0.00295)	(0.00726)
Focal reader assigned label to previously read book(s) by the focal author	0.835***	0.770***
	(0.00262)	(0.00487)
Total labels applied to focal book by reader, ln	-0.0000437***	-0.0000662**
	(0.00000574)	(0.0000215)
Review year	0.000000489	0.00000255
	(0.000000816)	(0.00000276)
Constant	-0.00175	-0.00652
	(0.00165)	(0.00557)
User FE, Book FE, Label FE	Y	Y
N	20101544	2356722
Adj. R-sq	0.795	0.729
Log likelihood	25886346.8	2711091.7

Standard errors (clustered on users, books, and labels) in parentheses
* p<0.05, **p<0.01, ***p<0.001

Table A6 Individual-level estimates of label mismatch

	(1)	(2)	(3)	(4)
DV: Individual-level label mismatch	Full sample		Matched sample	
Individual-level extent of repositioning	0.0368***	0.0189***	0.0430***	0.00583
	(0.00200)	(0.00354)	(0.0110)	(0.0214)
Individual-level extent of repositioning -- squared		0.0269***		0.0571
		(0.00579)		(0.0336)
Count of author's books rated by individual (in thousands)	-0.188*	-0.136		
	(0.0802)	(0.0808)		
Total labels applied to focal book by individual (ln)	-0.351***	-0.351***	-0.357***	-0.357***
	(0.00198)	(0.00198)	(0.00378)	(0.00377)
Review year	0.000220	0.000228	-0.00114*	-0.00113*
	(0.000143)	(0.000148)	(0.000521)	(0.000521)
Constant	0.368	0.351	3.101**	3.079**
	(0.288)	(0.298)	(1.049)	(1.049)
User FE	Y	Y	Y	Y
Book FE	Y	Y	Y	Y
N	1285992	1285992	64627	64627
Adj. R-sq	0.615	0.615	0.615	0.615
Log likelihood	600854.8	600890.0	42897.2	42902.0

Standard errors (clustered on users and books) in parentheses
* p<0.05, **p<0.01, ***p<0.001
Note: While the linear and quadratic terms of individual level repositioning in Model 4 are not individually significant, they are jointly significant, i.e., they improve model fit significantly according to a LR test (p<0.01).

Table A7 Book-level estimates of effect of author's diversity in prior books on label mismatch

DV: Label mismatch	
Extent of repositioning	0.323***
	(0.0109)
Author's diversity in prior books	-0.594**
	(0.185)
Extent of repositioning X	-9.632***
Author's diversity in prior books	(0.435)
Total labels applied to focal book, ln	-0.0305***
	(0.000575)
Author book order (ln)	-0.00607***
	(0.00167)
Publication year of focal book	-0.00321***
	(0.000149)
Constant	6.943***
	(0.297)
Author FE	Y
N	255106
adj. R-sq	0.250
Log likelihood	105017.9

Standard errors (clustered on authors) in parentheses
* p<0.05, **p<0.01, ***p<0.001
Note: Sample consists of authors' third and later books.

Table A8: Book-level estimates of total Goodreads ratings counts and mean rating, matched sample, continuous measure of label mismatch

	(1) DV: label mismatch	(2) DV: Total count of ratings (ln)	(3)	(4) DV: Count of 5-star ratings (ln)	(5)
Label mismatch (continuous)			-0.628***		-0.666***
			(0.0314)		(0.0338)
Extent of repositioning	0.113***	-0.433***	-0.362***	-0.440***	-0.364***
	(0.0136)	(0.0309)	(0.0343)	(0.0282)	(0.0303)
Repeat readership	-0.00794**	-1.596***	-1.601***	-1.363***	-1.368***
	(0.00299)	(0.0198)	(0.0193)	(0.0229)	(0.0223)
Constant	0.408***	4.242***	4.499***	2.930***	3.202***
	(0.00492)	(0.017)	(0.023)	(0.0172)	(0.0235)
N	262503	262503	262503	262503	262503
Adj. R-sq	0.013	0.153	0.16	0.112	0.12
Log likelihood	54590.5	-452078.4	-450984.2	-458940.4	-457770.2

Standard errors (clustered on matched strata) in parentheses
* p<0.05, **p<0.01, ***p<0.001

Table A9: Book-level estimates of total Goodreads ratings counts and mean rating, matched sample, label mismatch variable split into terciles

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	DV: Total count of ratings (ln)			DV: Count of 5-star ratings (ln)			DV: mean rating		
Label mismatch (high)	-0.272***		-0.261***	-0.296***		-0.285***	-0.0619***		-0.0622***
	(0.023)		(0.0216)	(0.0237)		(0.0225)	(0.00746)		(0.00761)
Repeat readership	-1.588***	-1.578***	-1.589***	-1.383***	-1.373***	-1.384***	0.0333***	0.0393***	0.0335***
	(0.026)	(0.0265)	(0.026)	(0.0288)	(0.0295)	(0.0289)	(0.00977)	(0.0097)	(0.00962)
Extent of repositioning		-0.406***	-0.368***		-0.409***	-0.366***		0.00259	0.0109
		(0.037)	(0.0452)		(0.036)	(0.0422)		(0.0196)	(0.0184)
Total count of ratings (ln)							0.0235***	0.0257***	0.0236***
							(0.00236)	(0.00245)	(0.00233)
Constant	4.255***	4.122***	4.301***	2.952***	2.802***	2.997***	3.741***	3.689***	3.740***
	(0.0208)	(0.0231)	(0.0233)	(0.0218)	(0.0237)	(0.0237)	(0.0112	(0.0127)	(0.011)
N	151131	151131	151131	151131	151131	151131	151131	151131	151131
Adj. R-sq	0.176	0.171	0.178	0.135	0.129	0.137	0.009	0.005	0.009
Log likelihood	-253032.1	-253448.9	-252783.6	-257528	-258041.7	-257295.1	-97220.2	-97511.5	-97218.5

Standard errors (clustered on matched strata) in parentheses
* p<0.05; ** p<0.01; *** p<0.001
Notes: Treatment is based on above- versus below-median values of label mismatch. The sample is constructed by starting with all books from all authors who have written two or more books and creating a matched sample that pairs each book in the highest tercile of label mismatch with a book that is in the lowest tercile of label mismatch but is observationally equivalent in terms of the following metrics: deciles of the total count of ratings of all prior books by the author, deciles of mean rating of all prior books by the author, deciles of the focal book's publication year, exact count of prior books published by the author, and the most common predicted genre label of the focal book. (Books in the middle tercile of label mismatch are dropped from this analysis.)

Table A10: Book-level estimates of total Goodreads ratings counts and mean rating, matched sample, splitting new versus repeat readers

	(1)	(2)	(3)	(4)	(5)	(6)
	DV: Total count of ratings (ln)		DV: Count of 5-star ratings (ln)		DV: Mean rating	
	New readers	Repeat readers	New readers	Repeat readers	New readers	Repeat readers
Label mismatch (high)	0.0197	-0.105***	-0.0426*	-0.124***	-0.0333***	-0.0452***
	(0.0256)	(0.0128)	(0.0181)	(0.0123)	(0.00389)	(0.00596)
Extent of repositioning	-0.221***	-0.262***	-0.253***	-0.234***	-0.0147	0.0139
	(0.0663)	(0.0288)	(0.043)	(0.0275)	(0.0117)	(0.0166)
Repeat readership					0.406***	-0.134***
					(0.0115)	(0.0086)
Total count of ratings (ln)					0.0274***	0.00919***
					(0.00141)	(0.00164)
Constant	2.583***	2.448***	1.723***	1.464***	3.620***	3.949***
	(0.0355)	(0.0128)	(0.0238)	(0.012)	(0.00865)	(0.0102)
N	262503	262503	262503	262503	193554	262503
adj. R-sq	0.001	0.003	0.001	0.004	0.047	0.009
Log Likelihood	-545374.9	-449314.6	-492137.9	-427288.6	-115034.5	-237619.6

Standard errors (clustered on matched strata) in parentheses.
* p<0.05; ** p<0.01; *** p<0.001
Formal tests of coefficient differences in models 1 vs. 2, 3 vs. 4, and 5 vs. 6 confirm that the coefficients on label mismatch are stronger for repeat readers.

Table A11 Individual-level estimates of effect of reader omnivorism

DV: individual rating	(1)	(2)	(3)	(4)
Individual-level label mismatch	-0.0122***	-0.0353***	-0.0347***	
	(0.000619)	(0.00307)	(0.00266)	
Individual omnivorousness: all books		-9.795***		
		(0.0412)		
Individual-level label mismatch X Individual omnivorousness: all books		0.463***		
		(0.0846)		
Individual omnivorousness: focal author's books			0.313***	7.026***
			(0.0611)	(0.0822)
Individual-level label mismatch X Individual omnivorousness: focal author's books			0.957***	1.634***
			(0.146)	(0.197)
Review year	0.000366***	0.00561***	-0.000515***	-0.00810***
	(0.0000447)	(0.0000475)	(0.0000737)	(0.0000752)
Constant	3.213***	-7.007***	5.099***	20.25***
	(0.0901)	(0.0953)	(0.148)	(0.152)
User FE	Y	Y	Y	Y
Book FE	N	N	N	Y
N	78,603,010	77,065,853	27,368,254	27,332,766
Adj. R-sq	0.211	0.212	0.281	0.376
Log likelihood	-100189838.0	-98140020.4	-31597823.0	-29507852.4

Standard errors (clustered on matched strata) in parentheses.
* p<0.05, ** p<0.01, *** p<0.001
NOTE: table includes all individual ratings, not only the ones where individual level labels are available. The number of observations drop slightly from model 1 to model 2 because the individual omnivorousness variable is not defined for the first two books read by the reader.

Table A12: Book-level estimates of total Goodreads ratings counts, matched sample, negative binomial models (robustness checks to Table 5)

	(1)	(2)	(3)	(4)	(5)	(6)
	DV: Rating count			DV: 5-star rating count		
Label mismatch (high)	-0.191***		-0.182***	-0.237***		-0.228***
	(0.0482)		(0.0449)	(0.0594)		(0.0552)
Repeat readership	-2.278***	-2.257***	-2.280***	-2.202***	-2.178***	-2.207***
	(0.109)	(0.0955)	(0.102)	(0.124)	(0.107)	(0.115)
Extent of repositioning		-0.770***	-0.760***		-0.892***	-0.882***
		(0.0889)	(0.0959)		(0.107)	(0.114)
Constant	6.121***	6.095***	6.217***	5.041***	5.001***	5.152***
	(0.0638)	(0.0569)	(0.0605)	(0.0762)	(0.0612)	(0.0714)
Ln alpha	0.887***	0.884***	0.882***	1.136***	1.132***	1.129***
	(0.0137)	(0.0126)	(0.0131)	(0.0139)	(0.0129)	(0.0133)
N	262503	262503	262503	262503	262503	262503
Log likelihood	-2054042.3	-2053065.5	-2052476.4	-1581960.4	-1581075.9	-1580358.1

Standard errors (clustered on matched strata) in parentheses.
* p<0.05; ** p<0.01; *** p<0.001

Figure A1. Example of the survey prompt used in the Prolific survey validating the results of the deep learning model

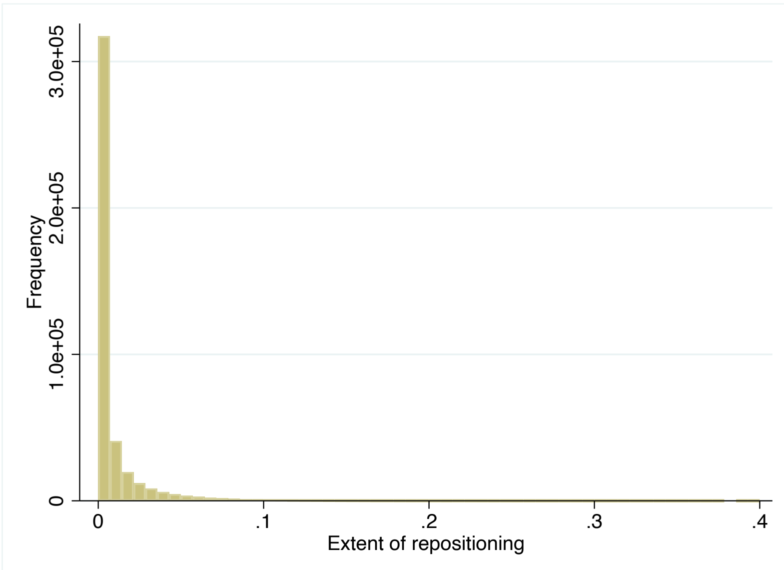
Here is the short description of a book. How would you tag this book? Please select all the tags that you think apply.

Ruthie loves Superman. Ruthie wants to be Superman. And when Ruthie is asked to go spend the afternoon with her aunt, who is about to have a baby any day day now and may need some help., Ruthie seizes the opportunity. It could be her chance to be a hero, should the baby come while she's visiting! But when Ruthie is out fetching a snack for her aunt, she gets so distracted by a box full of kittens in the bodega that she doesn't hear her aunt calling for her, nor does she notice the policemen running to the apartment or the ambulance pulling to the curb. When she realizes what's happened, she's devastated -- she's missed her one chance to be a hero! Or has she? Sonia Manzano, best known as "Maria" on Sesame Street, once again captures the warmth, love, and adventures of her childhood Bronx neighborhood.

☐ Biography	☐ Fantasy	☐ Music	☐ Science
☐ Business	☐ Fiction	☐ Mystery	☐ Science Fiction
☐ Chick Lit	☐ Gay and Lesbian	☐ Nonfiction	☐ Self Help
☐ Children's	☐ Graphic Novels	☐ Paranormal	☐ Spirituality
☐ Christian	☐ Historical Fiction	☐ Philosophy	☐ Sports
☐ Classics	☐ History	☐ Poetry	☐ Suspense
☐ Comics	☐ Horror	☐ Psychology	☐ Thriller
☐ Contemporary	☐ Humor and Comedy	☐ Religion	☐ Travel
☐ Cookbooks	☐ Manga	☐ Romance	☐ Young Adult
☐ Crime	☐ Memoir		

→

Figure A2. The distribution of the "extent of repositioning" variable (books' claimed content-based distance from previous book(s) of the author)



APPENDIX B

Predicted label proportions, label distance, superordinate vs lower level categories and category contrast

The predicted label proportions outputted by the deep learning model naturally capture several properties of label spaces documented in the prior literature: distances between labels, differing levels of abstraction, and category fuzziness. First, prior literature (e.g., Kovács and Hannan 2015) argued that distances between labels need to be considered when measuring category positioning. Deep learning, and BERT specifically, naturally captures label distances. In BERT, labels are represented as positions in the 768-dimensional space, while books are also represented as positions in the same space. Predicted label probabilities are calculated as a decreasing function of distance between the label's position and the book's position. And because the BERT representation is a metric space, labels close in the 768-dimensional space (i.e., similar labels) will also have similar predicted probabilities for a given book. Overall, the predicted probabilities of similar labels will be positively correlated. For example, the predicted probabilities for "war" and "history" correlate at 0.70; "crime" and "mystery" correlate at 0.75. The predicted probabilities of dissimilar labels will be negatively correlated (e.g., "autobiography" and "fantasy" correlate at -0.13). Because distances are already captured by BERT, no additional label-distance adjustments, such as the Jaccard-distance used in Kovács and Hannan (2015), are needed.

Second, existing research has argued that categories could be construed at different levels of abstraction and organized to superordinate and subordinate categories (e.g., Mervis and Crisafi 1982). Younkin and Kashkooli (2020) recently argued that taking the level of abstraction into account matters when investigating category positioning because audiences may resolve contentious categorizations using superordinate categories. This point is also crucial for our paper because an author's new book may be construed as repositioning at one level of abstraction (thriller -> mystery) but not at the superordinate level (fiction -> fiction). Relatedly, the extent of misclassification may depend on the level of category abstraction.

The level of abstraction of the labels Goodreads readers attach to books do vary[16], from the high-level "fiction" to the more specific "romantic suspense." How does such variance across the abstraction level of labels affect our measures of misclassification and repositioning? By construction, BERT takes the generality of labels into account when calculating distances. More general labels will be associated with a higher number of different labels, and their position in the 768-dimensional space will be in between the positions of the labels they are associated with (for example, fiction will be in between romance, historical fiction, etc.). This means that repositioning between lower-level labels such as romance to sci-fi will show up as a larger distance than repositioning from a lower-level to a higher-level label (from romance to fiction) or between two higher-level labels (fiction to fiction, i.e., 0 distance). In the Appendix, we show additional results in which we remove superordinate-level labels such as fiction and still find results consistent with our hypotheses.

Finally, research on category positioning and specialization has demonstrated that the contrast of the categories spanned influences audience reactions (e.g., Kovacs and Hannan 2010). While we do not model category contrast directly, we note that our distance-based measure of repositioning captures the effect category contrast may have on the effect of repositioning. Namely, labels that are sharp tend not to occur together with other labels. Thus, repositioning between sharp labels will be captured as a large distance spanned, while repositioning between fuzzy labels will be perceived as a small distance spanned.

[16] Crowd-sourced labels such as used in Goodreads typically cannot be neatly organized into hierarchical trees with super- and supraordinal categories. First, labels can belong to multiple higher level labels, for example "war" is associated with both fiction and non-fiction. Second, a hierarchical clustering would require asymmetric relationships such that if A belongs to B then B doesn't belong to A. This assumption is violated in the Goodreads data, for example not all chicklit books are fiction and not all fiction are chicklit. Overall, we think that the conceptual space of Goodreads labels is closer to an associative network than to a hierarchical classification.