

Facets of Specialization and Its Relation to Career Success: An Analysis of U.S. Sociology, 1980 to 2015

American Sociological Review
2021, Vol. 86(6) 1164–1192
© American Sociological
Association 2021



DOI:10.1177/00031224211056267
journals.sagepub.com/home/asr



Raphael H. Heiberger,^a 
Sebastian Munoz-Najar Galvez,^b 
and Daniel A. McFarland^c 

Abstract

We investigate how sociology students garner recognition from niche field audiences through specialization. Our dataset comprises over 80,000 sociology-related dissertations completed at U.S. universities, as well as data on graduates' pursuant publications. We analyze different facets of how students specialize—topic choice, focus, novelty, and consistency. To measure specialization types within a consistent methodological frame, we utilize structural topic modeling. These measures capture specialization strategies used at an early career stage. We connect them to a crucial long-term outcome in academia: becoming an advisor. Event-history models reveal that specific topic choices and novel combinations exhibit a positive influence, whereas focused theses make no substantial difference. In particular, theses related to the cultural turn, methods, or race are tied to academic careers that lead to mentorship. Thematic consistency of students' publication track also has a strong positive effect on the chances of becoming an advisor. Yet, there are diminishing returns to consistency for highly productive scholars, adding important nuance to the well-known imperative of publish or perish in academic careers.

Keywords

field change, sociology of science, scientific careers, structural topic modeling

Sociology operates on the verge of its own undoing, seemingly without respite. Throughout the twentieth century, many scholars sounded the alarm around the fragile state of the discipline: sociology is disintegrating (Znaniecki 1945:514), amid a clash of theories (Giddens 1976:715), compartmentalized into “congeries of outsiders to each other” (Collins 1986:1340–41), with subspecialties drifting apart from each other (Hand and Judkins 1999:18). As early as 1920, we find calls to map the field (Committee on Standardization of Research 1920, cited in Rhoades 1981:16). It seems the discipline can only be

pursued, if at all, against its own tendency for self-subversion (Holmwood 2010:650).

Yet, after a century of sociology's alleged disintegration, students continue to pursue the discipline. In fact, the number of awarded

^aUniversity of Stuttgart

^bHarvard University

^cStanford University

Corresponding Author:

Raphael H. Heiberger, University of Stuttgart,
Institute for Social Sciences, Seidenstr. 36,
Stuttgart, 70174, Germany
Email: raphael.heiberger@sowi.uni-stuttgart.de

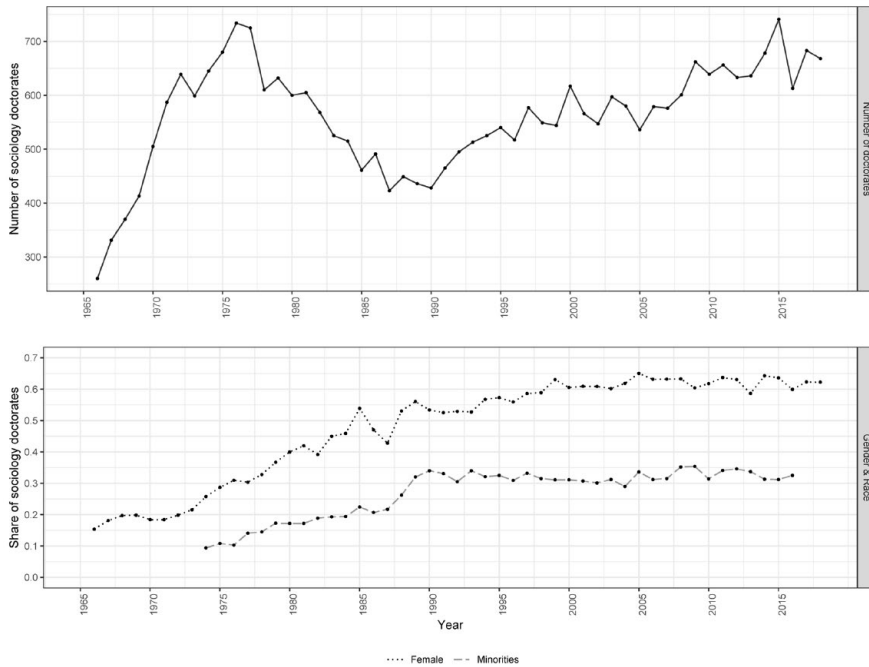


Figure 1. Demographic Developments in U.S. Sociology

Note: Development of total number of awarded doctorates in sociology in the United States (upper panel). Data on the number of students is taken from the Survey of Earned Doctorates (National Science Foundation 2020). In the lower panel we depict the compositional change of doctorate recipients in sociology by gender and race. Definition of students' gender and race resembles the methodology used by the ASA (2021b, 2021c).

PhDs in sociology has increased steadily since the 1990s, and in recent years, it is close to the prior peak of the late 1970s (see Figure 1), which is known as a period of critical overcrowding (Collins 1986:1338; Crane and Small 1992:199). The field's demographics changed too. Since the late-1980s, women have been awarded the majority of sociology doctorates, and the share of ethnic minority graduates climbed from approximately 10 percent in the 1970s to around one third in 2016. In contrast, the number of faculty positions has slightly declined since 2000 (ASA 2021a). These antagonistic trends might increase the pressure to publish early in one's career, which, in turn, raises the bar that young scholars must clear to secure a permanent position in sociology (Warren 2019).

Specialization has long been recognized (Collins 1986; Weeber 2006), maligned (Coser

1990; Gitlin 1990), and studied (Leahey, Keith, and Crockett 2010; Leahey and Moody 2014) as a publication strategy of extant sociology faculty, but little attention is given to how graduate students adopt this strategy. This is surprising, given that young scholars are charged with reproducing and transforming the field (Bourdieu 1988:94; Kuhn 1962:151). Our objective here is therefore to answer, how do sociology students specialize and to what professional returns? To observe students' specialization, we utilize an understudied genre of writing: the dissertation, which, we argue, enacts a student's appeal for recognition in one of the field's research areas.

The academic specialization process entails adopting a particular research focus, and sustaining this focus consistently over time to become a recognized expert. This focus can vary in degree and targeted content.

It also occurs within a larger context of available themes in the field, with respect to which it can be described as novel or traditional specialization. We synthesize these facets of specialization into three antinomies: tradition versus innovation (Foster, Rzhetsky, and Evans 2015; Kuhn 1959; Leahey and Moody 2014), focus versus breadth (Leahey 2007; Leahey et al. 2010; Teodoridis, Bikard, and Vakili 2019), and consistency versus change (Koppman and Leahey 2019; Trapido 2015). With regard to student dissertations, the first two facets concern students' topical choices: Does their thesis deal with a narrow set of topics? Is it within an identifiable research tradition, or does it span across them? The last facet concerns the extended publication track of students after graduation: Do they continue to publish in the same thematic area as their dissertation or do they change their focus? In addition to these three facets, we also consider specialization in the sense of students' choice of specific topics, which we call targeted specialization.

To study research specialization, we focus on the language used in student theses and ensuing published works. We fit a structural topic model (STM) (Roberts, Stewart, and Airolidi 2016; Roberts et al. 2014) to the abstracts of student dissertations. We recover topics, or "word systems" (Astley 1985:499), reflective of the research areas directly expressed in young scholars' writings, and then use these topics to develop measures for each facet of specialization. By linking this information to students' ensuing careers, we are able to ascertain the relation that different forms of specialization have on an important long-term career outcome in academia: time to becoming a primary advisor. Such attainment is a conservative proxy for success—only around 3 percent of PhDs in our broad sample become advisors of sociology-related theses—but it reflects if and when a student occupies an influential position in field reproduction. Furthermore, by focusing on cohorts of graduate students, our approach allows us to illuminate specialization strategies used at the start of a career, whereas prior studies

concentrating on salaries (Leahey 2007) or citations of extant faculty (Sinatra et al. 2016) may suffer from survival bias.

We connect facets of specialization to the time it takes until first mentorship via event-history models. We find students in 1980 to 2015 became advisors of sociology-related PhDs via distinct specialization strategies. Students were most likely to achieve influential field positions when they targeted certain topics reflective of the cultural turn or statistical models rather than adopting a thematic focus in general. They were also likely to secure such positions when they adopted novel topic-combinations rather than traditional forms. Finally, students were more likely to become advisors when they established a consistent research identity early in their careers. Such consistency need not be large, however. Only a small number of publications consistent with the dissertation provides considerably better chances to achieve a position as an advisor. This finding adds important nuance to the well-known imperative of publish or perish.

SPECIALIZATION AS CAREER STRATEGY IN SOCIOLOGY

During the plenary sessions of the 1988 annual meeting of the American Sociological Association, Gitlin (1990:216) argued that sociology had lost public value because students with ever more specialized interests invested in obscure scholastic writing rather than engage in debates of public import. Coser (1990:208–9) added that in this kind of highly specialized writing, the "methodological tail wags the substantial dog," and concepts lack critical bite. Moore's (1990) rejoinder in this debate argued, on the contrary, that African American, Latinx, and women sociologists, in particular, play a significant, albeit specialized, role in advancing the discipline's critical tradition, often in their teaching and in organizing for regional professional associations.

The 1988 ASA polemic occurred during a period in which young sociologists adopted

high degrees of specialization to avoid downward mobility into contingent academic work (Weeber 2006:59). More formally, in the 1980s and 1990s, young sociologists pursued specialization as an effective adaptation to an overcrowded academic market, in which they could only compete for the attention of their peers within narrow disciplinary niches (Collins 1986).

The narrative of fragmentation alluded to in the introductory paragraph, and echoed by Gitlin and Coser, is grounded in the true state of affairs of a growing field with many loosely connected themes and an impoverished corps—a discipline that is increasingly subject to the pressures of specialization and market-oriented higher education (Holmwood 2010). What this narrative obscures is that fragmentation is an emergent property of the way novice scholars vie for a place in the field; namely, through specialization. It is no surprise then, that previous efforts to investigate sociology—as a field (Leahey and Moody 2014) and career (Koppman and Leahey 2019; Leahey 2007; Leahey et al. 2010)—focus on the issue of specialization.

Our own contribution to a better understanding of specialization as a career strategy in sociology is threefold: we investigate an understudied site for specialization—the dissertation; a critical career outcome—becoming an advisor; and a suite of specialization strategies—topic choice, focus, novelty, and consistency.

The Dissertation Genre

The dissertation is one of the most decisive documents in a scholar's life.¹ It can be thought of as a claim to a career and to a research program worth pursuing (Bourdieu 1988:94). A thesis can also be thought of as a rite of passage (Collins 2002:43), whereby a young scholar shows deference to a field's canon (Paré, Starke-Meyerring, and McAlpine 2011:220). Of course, young scholars have also been described as central agents of field transformation (Kuhn 1962:151), a role that may also be expressed in the dissertation genre.

Dissertations constitute a site for the specialization strategies of young sociologists; this is the field constituency variously charged with reproducing or transforming the field. Also, this constituency's efforts to capture the attention of a niche audience fuel the disciplinary lament of fragmentation. In addition, unlike any other genre of academic writing, the dissertation does not suffer from publication bias, a pressing issue in the sociology of science (Fortunato et al. 2018).

Becoming an Advisor

Becoming an advisor encodes a crucial aspect of the success of a specialization strategy. It is through relations of mentorship that scholars recruit allies into intellectual movements and legitimate novel academic identities (Frickel and Gross 2005). Which graduates do or do not become advisors, and the time it takes to become one, reflects which forms of specialization are in most immediate demand. Scholars who achieve this stage become gatekeepers of the specialized disciplinary niche. Also, intergenerational networks of mentor-mentee relationships can reinforce the status of different methodologies in sociology subfields (Bandelj 2019). These are also central to the transmission of highly codified research practices (Leahey 2006).

Despite the critical importance of advisorship as a career achievement, particularly in relation to specialization strategies, this is not an outcome that receives much attention in the literature. Many scholars emphasize rewards granted to esteemed faculty as important points in academic careers—for instance, prestigious prizes (Zuckermann 1977), higher salaries (Leahey 2007), or high-impact papers (Sinatra et al. 2016). Just a handful of studies focus on the acquisition of long-term positions, like (tenured) professorships (Balsmeier and Pellens 2014; Leahey et al. 2010; Lutter and Schröder 2016). This work finds that besides the quantity and quality of the basic unit of scholarly achievement (i.e., publications), other non-meritocratic factors, especially gender (van den Besselaar

and Sandström 2016) and affiliations at elite institutes (Burris 2004), influence the chances of becoming faculty. Instead of using appointments to professorships or getting tenure as a proxy for academic success, we focus on all graduates of sociology-related theses and identify which go on to become advisors of sociology-related theses themselves and the time to such a position. We argue this is a more conservative and substantive approximation of achieving an influential and successful scientific career.

There is a considerable gap between the site in which we will study specialization—the dissertation—and the career stage in which we will measure the success of this strategy—becoming an advisor. To connect the two, we conceptualize and measure different facets of specialization, some of them narrowly connected to the document of the dissertation; others extending outward from the dissertation to a scholar's entire documented research output.

FACETS OF SPECIALIZATION

Students write about specific methodologies, theories, and subject matters that appeal to distinct audiences in the field. Through the selection of topics, students can display varying degrees of thematic latitude (focus versus breadth); furthermore, they can consolidate prior links and traditions, or combine topics that have rarely been combined (tradition versus novelty). In their future work, students can persist in the same vein of their dissertation or venture out to other themes (consistency versus change). These different facets of specialization have rarely been studied together, or with an approach that allows the specialized subject matter to emerge directly from students' writing.

Specializing on Target Topics

A dissertation draws from multiple topics; that is, multiple methods, theories, and subject matters expressed in careful word choices that elicit recognition from peers. This community

of peers may be organized in specialized sections of a professional organization (Leahey 2007; Leahey and Moody 2014); or it may be more loosely structured as a collective that shares a set of methods and ontologies (Knorr Cetina 2009) and "similar views on contested scientific topics" (Teplitskiy et al. 2018:1828). In either case, the topic structure expressed in dissertation writing displays students' aptitude to elicit their community's signature thought style (Antons, Joshi, and Salge 2019:3044–5; Fleck 1979). This crucial act of position-taking through topic choice will have differential professional returns, as students will vary in their success dealing with a topic, and the communities students appeal to through their topic choices may have more or less to offer them in terms of career prospects.

The effect of *targeted specialization* in specific topics on career outcomes is not well understood (Gieryn 1978). Among the few studies that exist, Hoppe and colleagues (2019) find that in medicine, selecting certain topics contributes to lower success rates of African American scientists at the NIH. Munoz-Najar Galvez, Heiberger, and McFarland (2020) find that in education, targeting topics associated with the interpretive approach improves PhD students' chances of an academic career. And Moody (2004) finds that for sociology, specialty areas shape collaboration networks, and quantitative work is more likely to be coauthored. Based on the extant literature, we formulate a baseline hypothesis that specializing more than others in specific topics will likely shape career prospects:

Hypothesis 0: Targeting certain topics will affect PhD students' chances of becoming an advisor.

Focus versus Breadth

In contrast to the effect of targeted specialization, much effort has been put into disentangling whether specialists or generalists have more successful careers. This distinction captures the trade-off between spending limited

resources to gain an in-depth understanding of a specific domain versus gathering more superficial knowledge across several domains. Previous research describes affordances and penalties for both strategies. For instance, specialized scholars who master the literature of a subfield can leverage this mastery to write more papers (Leahey 2007) or produce breakthrough inventions (Conti, Gambardella, and Mariani 2014). In the sociology of concepts and categories, Hannan and colleagues (2019:4) argue that, in a variety of contexts, “a clear position in one category tends to increase an offer’s appeal”; they call the obverse of this affordance the “category-spanning discount.”

In contrast, Teodoridis and colleagues (2019) find that affordances for both specialist and generalist strategies apply in creative work, depending on the pace of change in a given field, with generalists being the more desirable collaborators in slower fields. A related stream of work argues that creative workers should seek boundary-spanning combinations that achieve greater recognition (Burt 2004; Fleming 2001; Fleming, Mingo, and Chen 2007; Schilling and Green 2011). Acknowledging the competing evidence, as well as the lack of pertinent case studies dealing with the dissertation genre or academic field novices, we hypothesize two competing claims:²

Hypothesis 1a: A topically focused thesis leads to higher chances of becoming an advisor.

Hypothesis 1b: A topically broad thesis leads to higher chances of becoming an advisor.

Novelty versus Tradition

Several studies find that novel combinations of prior knowledge increase the recognition of academic papers. This line of work describes unusual combinations of chemical substances (Foster et al. 2015), bibliographic references (Uzzi et al. 2013), disciplines (Leahey, Beckman, and Stanko 2017), subfields (Leahey and Moody 2014), concepts (Hofstra et al. 2020), and topics (Antons

et al. 2019) as garnering outsized recognition. However, there is also evidence that papers benefit from closeness to established knowledge. Foster and colleagues (2015) make clear that the “bread and butter” business of academia (i.e., job talks and grants) rewards the production of traditional work, and Shi, Leskovec, and McFarland (2010) show that work squarely situated within and consistent with a research community has the likeliest chance of recognition. Fleming (2001), Uzzi and colleagues (2013), Trapido (2015), and Leahey and colleagues (2017) all point out that novel recombination of ideas and categories represents a high-risk high-reward endeavor that often fails, but sometimes yields great success.

It is unclear how novelty reflects on young scholars and their prospects to become advisors. Younger scholars are more likely to pursue novel research (Bourdieu 1988; Kuhn 1962), but they often face precarious job prospects. Thus, we posit competing claims between innovative and traditional combinations (Kuhn 1959):

Hypothesis 2a: Novel topic combinations in a thesis increase the odds of becoming an advisor.

Hypothesis 2b: Traditional topic combinations in a thesis increase the odds of becoming an advisor.

Consistency versus Change

Our hypotheses thus far connect crucial features of specialization in students’ dissertation writing to a long-term outcome of their academic careers: becoming an advisor. We reviewed a broad literature showing the relevance of specialization to career outcomes. But the link between an initial state of specialization and a long-term outcome relies on the assumption that entry conditions into an academic field can credibly affect how a student’s career unfolds. This link has been theorized in classical works in the sociology of science, where scholars’ initial standing in a field is said to express and reproduce stable

aspects of their institutional and social status (Bourdieu 1988). Of course, this assumption needs to be tested empirically. To do that, we introduce another facet of specialization, which we refer to as consistency versus change. This facet describes the link between the topic structure of a student's dissertation and topic choices in ensuing work, leading up to the career stage when the person may be considered to become an advisor.

Patterns of consistency (versus change) affect the unfolding of professional careers. For a new scholar, persisting in the same research area is a safe bet because changing specializations requires starting over, and chances of failure are high (Gieryn 1978:105–7). Yet, there are times when a significant pivot might pay off. For instance, movie actors increase their chances of being hired by playing similar roles early in their careers, and then by playing against type later on (Zuckerman et al. 2003). By contrast, Indian bureaucrats are rewarded for being a generalist early and adopting a more specialized and consistent identity later (Ferguson and Hasan 2013).

Closer to our case, Koppman and Leahey (2019) find that in sociology, long-standing commitments to a relevant identity, such as interdisciplinarity, increase a scholar's likelihood of using unconventional methods. The authors also find that affiliations to high-ranking universities add to the effect of a consistent identity; a result that is consistent with cumulative advantage (Correll et al. 2017). Similarly, for academic engineers, Trapido (2015) shows that audiences are more likely to deem authors' novel contributions valuable when they are consistent with previously developed identities as innovators. Synthesizing previous findings on the effects of consistency leads to a third pair of competing hypotheses:

Hypothesis 3a: A consistent research identity (persisting with topics expressed in one's thesis) improves the odds of becoming an advisor.

Hypothesis 3b: A changing research identity (departing from topics expressed in one's thesis) improves the odds of becoming an advisor.

ANALYTIC STRATEGY

To study how student theses reflect strategies of specialization, we draw on approximately all doctoral theses from 1980 to 2015 that conduct research related to sociology. We apply structural topic models (STM) (Roberts et al. 2016; Roberts et al. 2014) to those texts to identify the themes used in students' writings. In so doing, we adopt a methodological frame that uses the actual language in students' theses. The identified themes cut across fixed schemes of pre-labeled categories often used in prior work. We argue that students' specialization strategies are reflected in how they use these topics, and we develop metrics indicative of these strategies' usage. Finally, we use event-history models to test hypotheses as to which specialization strategies enable graduates to quickly reach sociology advising positions.³

Sample

Our analyses draw on a comprehensive database of doctoral theses written in the United States provided by ProQuest. We selected all dissertations in which the stem "sociol" was present in the title, abstract, keywords, or department ($N = 82,363$).⁴ In doing so, we cast a wide net for theses related to sociological "thought styles" (Fleck 1979), representing the disciplinary structure as fragmented and influenced by neighboring fields. This relatively broad definition allows us to trace graduates from multiple disciplinary backgrounds who also contribute to the sociological field (Lamont and Molnár 2002). Knowing the department of graduation also allows us to distinguish between students at sociology's "institutional heartland," that is, students graduating from sociology departments, and authors whom we label as sociology-related. The data comprise each dissertation's abstract, title, name of the graduate student, advisor's name, institutional affiliation, and National Research Council (NRC) discipline. In total, the records contain theses from 12 disciplines,⁵ cover 275

different U.S. institutions, and stretch from 1980 to 2015.

Identifying Thought Styles via Structural Topic Models

To discover thematic structures in dissertations, we use topic modeling (Blei 2012; Evans and Aceves 2016). Scholars have used topic models successfully to describe, for instance, the methodological divide of sociology (Schwemmer and Wieczorek 2020), themes of management research (Antons et al. 2019), math education (Inglis and Foster 2018), computer science (Anderson, McFarland, and Jurafsky 2012), and education research (Munoz-Najar Galvez et al. 2020). A conventional topic model takes as input a collection of texts formatted as a document-term matrix—a table in which every document is represented as a vector of word counts over the vocabulary of the document collection. The model learns the mixture of topics that make up each document and the words that belong to each topic via successive updates to two sets of prior multinomial distributions, (1) one of documents over topics and (2) one of topics over words.

We fit a structural topic model (STM) to the abstracts of sociology dissertations. STMs were developed by Roberts and colleagues (2014) to enable the use of document covariates (e.g., publication year) to improve the consistent estimation of topics. One of the ways an STM can include document covariates is in the estimation of a *document-specific* distribution over topics based on a document's covariates. In other words, in an STM, topic proportions can be modeled as a function of covariates.⁶ Our model uses a spline on year of publication to allow the average topic proportions to vary over time. We will describe the resulting trends in the first section of our results.

Like other unsupervised learning procedures (Molina and Garip 2019), however, researchers have to decide the number of topics (k) the model should identify. Note that this is problematic because the relevant topics

are not known a priori. We selected a solution with 60 topics after using a careful three-step approach to select k (see Part A of the online supplement).

Two output matrices are essential for us. The first contains a list of topics, each consisting of words ordered by their probability of being associated with the respective topic. Table 1 presents these topics using FREX values (Roberts et al. 2016), where a word is weighted by the harmonic mean of its rank in terms of frequency (FR) and exclusivity (EX) for a topic. For instance, among the most descriptive words for topic40 (T40) are “qualitative,” “theme_emerge,” “ground_theory,” “phenomenological,” and “live_experience.” It seems intuitive to assume theses with high loads of T40 are engaged in *qualitative methods*. Thus, labeling topics resembles qualitative coding approaches, because topics allow multiple interpretations depending on the perspective we use to look at them (Inglis and Foster 2018).

The second important output for our undertaking is the document-topic matrix, often denoted θ . This matrix contains each document summarized as a vector of the proportions of its words associated with each topic (Roberts et al. 2014:1069). These topic proportions represent the strength of association between each topic and a given dissertation. We use the θ matrix to construct metrics for students' topic choices, focus, novelty, and consistency.

Following Antons and colleagues (2019:3044–5), we argue that classification of scholarly work resulting from a topic model resembles Fleck's (1979) “thought styles” because it provides “a snapshot of the word choices representing the thought collective's encoding of its signature thought style.” Topics then “bear a more or less distinctive coloring conforming to a given thought style” (Fleck 1979:109), and we treat a dissertation described in terms of topics as a proxy for students' position in the sociological discourse—sociology's “word system” (Astley 1985:499)—at a crucial early career stage. These topics will later become a means to

Table 1. Topics in Sociological Dissertations, 1980 to 2015

ID	Label	FREX terms	Prevalence
1	<i>French/Can Soc</i>	french, france, canada, canadian, ghana, haitian, aboriginal, quebec	.006
2	<i>Identity</i>	identity, negotiate, discursive, construction, understanding, discourse, ethnography, everyday_life	.038
3	<i>Religion</i>	church, congregation, evangelical, denomination, pastor, worship, mormon, catholic	.010
4	<i>Race: Latinx</i>	mexican, mexican_american, los, angeles, puerto_rican, latina, chicano, puerto_rico	.008
5	<i>Self-perception</i>	perception, satisfaction, self_esteem, self_concept, self_perception, expectation, self, congruence	.019
6	<i>Violence</i>	violence, abuse, domestic_violence, victimization, rape, sexual_abuse, sexual_harassment, batter_woman	.011
7	<i>School</i>	teacher, public_school, school_district, teacher_student, charter_school, school_choice, teacher_perception, school_reform	.019
8	<i>Child Welfare</i>	foster_care, child_maltreatment, adoptive, adoptee, foster_child, adoptive_parent, permanency, child_welfare_system	.011
9	<i>Medical</i>	physician, hospitalization, mental_illness, medical, hmo, chronic_pain, medication, asthma	.012
10	<i>Rural/Environ</i>	forest, rural_community, climate_change, community, conservation, local_community, environmental_justice, earthquake	.016
11	<i>Method: Experiments</i>	experimental_group, treatment_group, pretest_posttest, posttest, program, intervention, program_evaluation, effectiveness	.018
12	<i>Econ Soc</i>	expenditure, income_inequality, household, saving, income, tax, estimate, cost	.025
13	<i>College Students</i>	freshman, grade_point_average, international_student, college_students, service_learn, campus, semester, gpa	.014
14	<i>Method: Survey/Scales</i>	score, subscale, locus_control, item, scale, anova, significant_difference_find, chi_square	.033
15	<i>Social Movement</i>	social_movement, protest, electoral, election, mobilization, activism, democracy, political	.025
16	<i>Internet</i>	internet, online, digital, gamble, technology, information, addict, gaming, facebook	.008
17	<i>Labor</i>	agricultural, farm, industrial, crop, irrigation, farmer, factory, capitalist	.018
18	<i>Gender: Sexuality</i>	gay, lesbian, homosexuality, gay_man, lgbt, transgender, sexual_identity, homophobia	.011
19	<i>Career</i>	career, mentor, career_development, career_choice, success, recruitment, professionalism, mentor_relationship	.012
20	<i>Participation</i>	participation, activity, recreation, volunteer, leisure, civic_engagement, camp, disability	.011
21	<i>Sociolinguistics</i>	linguistic, speaker, dialect, sociolinguistics, vowel, grammar, phonological, code_switch	.015
22	<i>Social Services</i>	service, board, agency, social_service, service_delivery, personnel, referral, fund	.016
23	<i>Status/Groups</i>	group, attribution, identification, group_member, cognitive, trait, preference, judgment	.021

(continued)

Table 1. (continued)

ID	Label	FREX terms	Prevalence
24	<i>Illness</i>	nurse, grief, nurse_home, hospice, bereavement, nurse_care, grieve, end_life	.006
25	<i>Theory</i>	philosophical, moral, habermas, rationality, theorist, weber, sociology_knowledge, epistemological	.034
26	<i>Science</i>	science, graph, algorithm, error, cluster, dna, parameter, simulation	.017
27	<i>Organizations</i>	collaboration, organizational_structure, stakeholder, participatory, nonprofit, organization, partnership, advocacy	.018
28	<i>Marriage</i>	couple, marital, intimacy, husband_wife, marital_quality, marriage, marital_adjustment, relationship_satisfaction	.012
29	<i>Gender: Feminist</i>	woman, infertility, feminist, gender_equality, abortion, patriarchy, woman_color, gender_relation	.011
30	<i>Ethnic Studies</i>	muslim, taiwanese, hindu, indigenous_people, arab, post_soviet, indigenous, ethnic_conflict	.013
31	<i>Sport/Consumption</i>	sport, tourism, athlete, athletic, brand, restaurant, football, student_athlete	.008
32	<i>Work</i>	work, employee, job_satisfaction, organizational_commitment, transformational_leadership, organizational_climate, emotional_intelligence, organizational_performance	.015
33	<i>Law</i>	law, legal, juror, attorney, legislation, lawyer, litigation, supreme_court	.016
34	<i>Criminology</i>	prison, inmate, juvenile, incarceration, offense, recidivism, prisoner, correctional	.013
35	<i>Fatherhood/Divorce</i>	father, stepfamily, single_mother, father_involvement, fatherhood, coparenting, single_parent_family, paternal_involvement, family_structure	.012
36	<i>Method: Survey/Opinion</i>	attitude, respondent, attitude_toward, orientation, survey, belief, acceptance, willingness	.025
37	<i>Arts</i>	music, theatre, cinema, jazz, film, hip_hop, artist, art	.013
38	<i>Demography</i>	fertility, earnings, labor_force_participation, contraceptive, employment, labor_market, birth_interval, job_search	.016
39	<i>Social Work</i>	social_work, therapist, trainee, social_work_education, msw, school_social_work, field_instructor, work_alliance	.009
40	<i>Method: Qualitative</i>	qualitative, theme_emerge, ground_theory, phenomenological, live_experience, interview, semi_structure_interview	.037
41	<i>Social Capital</i>	network, social_capital, social_network, trust, exchange, cooperative, interorganizational, network_structure	.017
42	<i>Higher Education</i>	faculty_member, ged, college_choice, graduate_school, professor, postsecondary_education, postsecondary, aspiration	.012
43	<i>Poverty</i>	homeless, homelessness, house, poverty, welfare_reform, homeownership, dog, relocation	.011
44	<i>Globalization</i>	global, south_africa, globalization, ngo, brazil, human_right, world_bank, cross_national	.022
45	<i>Conflict</i>	friendship, boy, conflict, behavior, aggression, mediation, aggressive, interaction	.014

(continued)

Table 1. (continued)

ID	Label	FREX terms	Prevalence
46	<i>Family Disorder</i>	sibling, family_function, alcoholic, adhd, sibling_relationship, family_environment, autism, eat_disorder	.012
47	Noise	fiction, victorian, poem, jewish_identity, literary, shakespeare, american_jewish, writer	–
48	<i>Adolescence</i>	adolescence, among_adolescent, teen, youth, suicidal_ideation, protective_factor, parental_monitor, african_american_adolescent	.017
49	<i>Parenting</i>	parent_involvement, parenting_style, parent_perception, head_start, school_readiness, parent, parent_child, parent_education_program	.012
50	<i>Method: Modeling</i>	model, contextual, theoretical_model, causal, explanation, hypothesis, indicator, individual_level	.043
51	<i>Urban</i>	neighborhood, police_officer, suburb, community_police, city, gentrification, gang, redevelopment	.012
52	<i>Aging</i>	elderly, retirement, old_adult, grandparent, grandchild, caregiver_burden, grandmother, old_people	.012
53	<i>Immigration</i>	immigrant, migrant, refugee, vietnamese, chinese_american, filipino, navajo, asian_indian	.017
54	<i>Motherhood</i>	maternal, toddler, maternal_depression, mother_infant, emotion_regulation, attachment, adult_attachment, mother_child_interaction	.011
55	<i>Stress</i>	cope, stressor, cope_strategy, stress, cope_behavior, hardiness, psychological_distress, veteran	.013
56	<i>Mass Media</i>	message, television, news, broadcast, news_medium, mass_medium, media, journalist	.011
57	<i>Knowledge</i>	knowledge, phase, plan, learn, strategy, approach, development, expert	.029
58	<i>Race: African American</i>	black, white, racial, racism, black_white, racial_identity, multiracial, race_relation	.015
59	<i>Health</i>	hiv, hiv_aid, obesity, health_behavior, nutrition, immune_deficiency, prenatal_care, bmi	.013
60	<i>Historical Soc</i>	change, society, shift, today, period, transformation, traditional, modern	.028

Note: We interviewed experts in various sociological research fields and asked them to derive labels for each topic based on the most probable and distinguishing terms of the STM. The “Prevalence” column represents a topic’s popularity, that is, the sum of its loads across all theses and years (in percent of all loads). Part A of the online supplement provides further information on why T47 is considered noise (see Figure A3 in the online supplement). We also provide an extended topic-term matrix including prob-scores and additional terms in the repository (<https://github.com/RapHei/Facets-of-specialization>).

measure facets of specialization (targeted specialization, focus, novelty, consistency) that we relate to students’ career outcomes.

Dependent Variable

Our dependent variable is the time to becoming a primary advisor of a sociology-related

PhD student. We view “sociology-related” theses as those that are defended in a sociology department or that contain the stem “sociol,” as described in the Sample subsection.⁷ The broad sample of dissertations related to sociology represents the permeability of disciplinary boundaries (Lamont and Molnár 2002). Our dependent variable, which

follows the same inclusion criteria as our sample, reflects a student's success at becoming a gatekeeper for future students aspiring to enter the sociological field.⁸ We interpret the time it took a student to first become an advisor as a conservative approximation of reaching an influential position in the sociological field.

To construct this variable, we matched all unique PhD student names to the names of primary advisors in our sample. This matching procedure captures students' transition from student to advisor within the U.S. graduate education system. Note that former students in our sample who move on to advise students on dissertations with no reference to sociology would not appear as advisors in our data.

Independent Variables

In this section, we describe the way we measure each facet of specialization: targeted specialization, degree of focus, degree of novelty, and consistency with ensuing published work. All our measures are built on top of the document-topic matrix θ , which represents documents as mixtures of topics.

A topic of targeted specialization is distinctive of a student's dissertation *relative to the extent to which peers use the same topic*. We want to capture that distinction among peers, because most topics (e.g., theory or methods) are used to a similar extent in many dissertations—that is, most topics in a dissertation cannot be discerned as targets of specialization. Therefore, we develop a metric that ascertains the extent to which students target a topic above and beyond most of their peers. Formally, the metric for targeted specialization measures if the topic load of a topic i in a document d ($\theta_{di,t}$) is larger than the 75th to 90th percentiles (in steps of 5) of that topic in the previous year across all theses ($\theta_{di,t} > \theta_{i,t-1}^{\text{percentiles}}$). This suite of threshold-based metrics captures targeted specialization with different levels of strictness. In the Results section, we depict only coefficients that are statistically significant across

all thresholds, which provides the most conservative estimates. Results for other thresholds are available in Part B of the online supplement.

The degree of *focus* in a thesis (Hypothesis 1) is measured by calculating the Herfindahl index for the discrete probability distribution over topics θ_d that summarizes each document d as a mixture of topics. Here we are following Ferguson and Hasan (2013), who used the Herfindahl index to measure the focus of prior experience in specific types of work and the effect on Indian administrative officers' careers. The Herfindahl index is calculated by summing the squared topic proportions over all topics k in a dissertation, that is, $\sum_i \theta_i^2$. This metric ranges from 0 to 1; larger values indicate a dissertation is more focused in a narrower set of topics.

Our measure of *novelty* (Hypothesis 2) captures the extent to which a dissertation introduces rare or unique *combinations of topics* with respect to prior dissertations. For that purpose, we adapt the relational measurement of “novel integration” used by Leahey and Moody (2014:236). We begin by converting the document-topic matrix, θ , into an occurrence matrix, where each dissertation receives a value of 1 on the two topics with highest proportions, and a value of 0 elsewhere. This matrix records the combination of the two most descriptive topics in each dissertation. Then, we subset this matrix to consider only the dissertations published three years prior to the document for which we want to measure novelty. Call this dissertation A, and imagine it combines the topics *higher education* and *violence*. We compute a ratio r of the number of times we observe the combination of *higher education* and *violence* to the expected occurrence of this combination. The novelty of dissertation A is calculated as $1 - r$, with higher values indicating greater novelty of its distinctive combination. We provide empirical examples in Part C of the online supplement.

Students' thematic *consistency* captures their continued focus on the topics first discussed in their dissertations. We gathered

each student's publications as part of a larger project in which we matched articles published by students in the ProQuest sample with papers recorded in the Web of Science database.⁹ In reconstructing students' publication records, we only consider articles in journals classified under the social sciences and humanities. In total, we match 140,746 articles to 18,548 students over a period of 35 years.

In a second step, we use our structural topic model to represent the abstracts of the journal publications as mixtures of the 60 topics that best describe U.S. dissertations related to sociology. We compare students' dissertation to their ensuing publications using cosine similarity as a measure of thematic distance—a conventional procedure in natural language processing (Manning, Raghavan, and Schütze 2008). These comparisons allow us to keep a running average of how consistent students remain with respect to the topics of their dissertation over time. In so doing, consistency provides a longitudinal measure to answer Hypothesis 3 on how coherent students' publication records appear in light of their dissertation, or, put differently, how consistent a research identity is when measured from the start of an academic career.

To close this section, we discuss the control variables of our models. We use three controls for academic standing: productivity, institutional prestige, and insider status. Productivity is the number of papers a student published, which we introduce to our models in log scale. For institutional prestige, we use an indicator variable that takes a value of 1 when the student's institution appears in the top 10 of the *U.S. News and World Report* rankings for graduate departments. This indicator is discipline and year specific when available. However, we had no recourse to comparable rankings before 1992, so we use that year as reference for all previous years.¹⁰ We also control for students' status as field insiders and outsiders. We define insiders as young scholars graduating from sociology departments; conversely, field outsiders are PhD students writing a thesis related to

sociology but receiving their doctoral degree from other departments.

We also control for students' *gender* and *race*, which are both covariates known to influence scientific careers (Cech et al. 2011; Huang et al. 2020; van den Besselaar and Sandström 2016). Gender has also been shown to influence the choice of research topics (Key and Sumner 2019), in general, and methods, in particular (Grant, Ward, and Rong 1987). Gender and race are derived from students' names by matching first names to Social Security Administration data to predict gender, and matching surnames to U.S. Census data to predict race (Hofstra et al. 2017; Hofstra and de Schipper 2018).¹¹ All our variables are described in Table 2.

Event-History Analysis

Our data provide information not only on who becomes an advisor of sociology PhDs, but how quickly they reach that position. On average, it takes 10 years between PhD and first serving as an advisor (see Part D of the online supplement). We take advantage of the time dependencies in our data by drawing on event-history analysis (EHA). This modeling strategy gives us two estimates of interest: the likelihood of becoming an advisor, and time to first advisorship.

EHA assumes a collection of units moves among a finite number of states that may occur at any point in time and are influenced by some observable factors (Coleman 1981:6). Eligible data should therefore include events, time elapsed until events happen, and time-constant or time-varying covariates. In our case, this setup models the time from PhD graduation to first becoming a PhD advisor, and how this process is influenced by facets of specialization reflected in one's thesis (topic choice, focus, novelty) and publication record (consistency), controlling for social attributes and performance (gender, race, prestige of alma mater, publications).¹²

The language of "hazard rate" of an event points to origins in health studies (Amorim and Cai 2015), but EHA has many applications

Table 2. Overview of Variables

Variable	Description	Operationalization	Mean	SD	Range
Advisor	Becoming an advisor	Dichotomous. PhD students who mentor advisees at sociology departments or advise theses that contain “sociol” (further details in notes 4 and 7).	.031	.174	[0,1]
Elite	Prestige of alma mater	Dichotomous. University of PhD students within top 10 of <i>US News</i> ranking in their department’s discipline.	.113	.316	[0,1]
Female	PhD student’s gender	Dichotomous. Estimated by using first name and Social Security Administration data (further details in note 11).	.545	.498	[0,1]
White	PhD student’s race	Dichotomous. Estimated by using full name and U.S. Census data (further details in note 11).	.708	.455	[0,1]
Insider	PhD at sociology or other department	Dichotomous. “Insider” (PhD granted by a sociology department) or “outsider” (PhD granted by other department).	.505	.500	[0,1]
Pub_Cum	Cumulated number of PhD student’s publications	Numeric. Number of publications of a PhD student as given in Web of Science (further details in note 9).	1.542	6.830	[0,359]
Targeted Specialization	Specialization in each topic relative to cohort	Dichotomous. 1 if the proportion of a topic in a thesis is higher than 75 to 90 percent of other students graduating in the previous year (H0). Details in Part B of the online supplement.	.101	.301	[0,1]
Focus	Focus versus breadth in topic proportions	Numeric. Degree of focus/breadth in a thesis calculated using the Herfindahl index (H1). Details in Part C of the online supplement.	.153	.077	[0,.951]
Novelty	Traditional versus novel topical combinations	Numeric. Degree a thesis integrates unusual topic combinations (H2). Details in Part C of the online supplement.	−.126	.949	[−8.27, 1]
Consistency	Consistent versus changing research identity	Numeric. Degree for how far or close students’ publications are in reference to one’s thesis (H3). Details in Part C of the online supplement.	.089	.213	[0,.99]

in social sciences and offers a variety of models (Blossfeld, Rohwer, and Schneider 2019; Box-Steffensmeier and Jones 2004; Singer and Willett 2003). We chose a parametric model with Weibull distribution to measure the time until first advisorship (see model choices in Part D of the online supplement).

RESULTS

Popular and Trending Topics in Sociological Dissertations

Between 1980 and 2015, the topics of sociological dissertations reflect a great diversity of subject matter subfields (e.g., T6 *violence*,

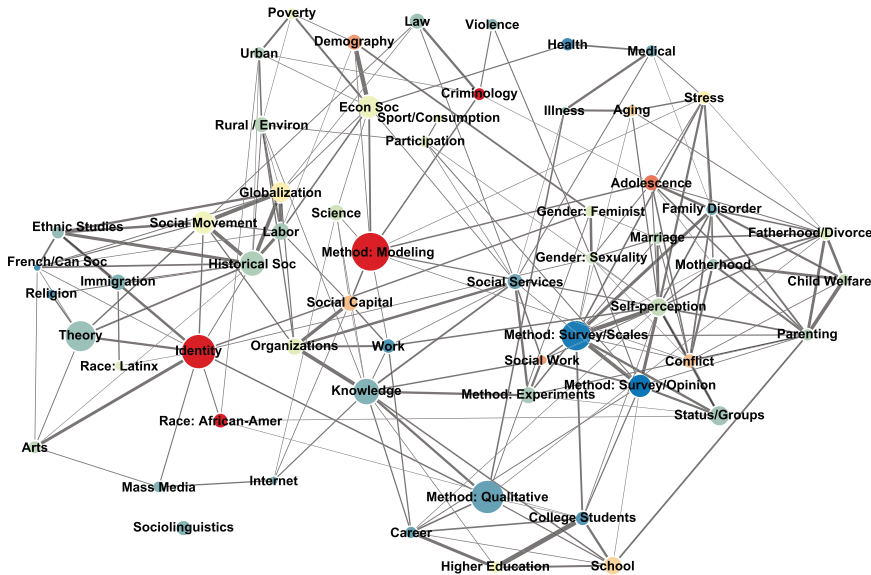


Figure 2. Network of Topic Co-usage in Sociological Dissertations, 1980 to 2015

Note: Topics connected by a tie and near each other indicate they are more likely to be discussed within the same thesis. The width of edges resembles the strength of correlation between topics given by their co-usage in dissertations. Only positive correlations are considered and a cutoff is set at the mean (.05) to facilitate readability. The size of nodes represents the relative topic prevalence over all years (see Table 2). The nodes are graphed using the ForceAtlas2 algorithm (Jacomy et al. 2014). The colors depict the likelihood of becoming an advisor if students specialize in this topic (i.e., targeted specializations as reported in Table 3, Model 5); shades of red depict higher likelihoods, blue lower likelihoods, and green represents effects close to zero.

T12 *economic sociology*, T34 *criminology*, T38 *demography*) and methods (e.g., T11 *experiments*, T14 *survey/scales*, T36 *survey/opinion*, T50 *modeling*, and T40 *qualitative methods*). Some topics reflect dimensions of social phenomena that span methods, theories, and empirics (T4 *race: LatinX*, T58 *race: African American*, T29 *gender*, T41 *social capital*). Finally, a group of topics comprise the vocabularies of neighboring disciplines (T5 *self-perception*, T21 *sociolinguistics*, T59 *health*).¹³

Some combinations of topics recur across many dissertations, which we can represent as a topic correlation network. Figure 2 shows the network of topic relations between 1980 and 2015, which reveals popular topic combinations. This network shows that methodology is a structuring principle of how students combine topics. For example, *qualitative methods* (T40) tends to be combined

with educational topics (T7 *school*, T13 *students*, T19 *careers*, T42 *higher education*, or T57 *knowledge*) as well as with T2 *identity* and T58 *race: African American*. As another example, we find that two closely connected topics about surveys—T36 *opinions* and T14 *scales*—link to themes related to family and social psychology.

The most prevalent topic—T50 *modeling*, with most distinctive terms “model,” “contextual,” “theoretical_model,” “causal,” “explanation”—occupies a bridging position in the topic network and is located between several thematic specialties (e.g., T12 *economic sociology*, T41 *social capital*, and T48 *adolescence*). To a lesser extent, T25 *theory* and T2 *identity* also exhibit bridging positions, yet both are more frequently linked to topics related to the cultural turn on the left-hand side of Figure 2 (e.g., topics on the study of ethnicity/race and the arts).

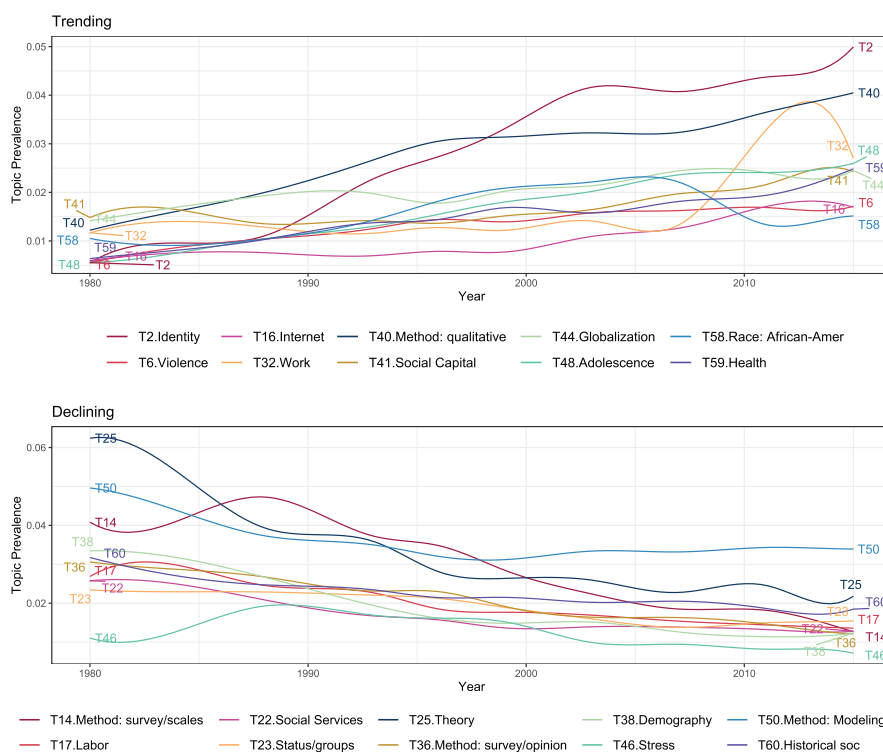


Figure 3. Most Pronounced Trends of Topics in Sociology, 1980 to 2015

Note: The 10 most trending (upper echelon) and declining (lower echelon) topics in U.S. theses related to sociology. Order is derived by using the slope of a linear regression of each topic's curve following Griffiths and Steyver's (2004) approach to capture hot and cold research topics. Topic prevalence describes the proportion of a topic's loading per year.

Figure 3 depicts the 10 most trending and declining topics in sociological theses during the period under study.¹⁴ The topic gaining most in students' popularity is T2 *identity*, which includes the following as its most distinctive terms: "identity," "negotiate," "discursive," "construction," and "understanding." T2 *identity* represents the intellectual movement often dubbed the "cultural turn" in social sciences (Alexander 1988; Jacobs and Spillman 2005).¹⁵ The cultural turn is concerned with processes "through which people define and express themselves" (Best 2007). Fittingly, T2 comprises concepts of discourse and meaning-making.

Almost parallel to T2 is the ascent of T40 *qualitative methods*. Among its most descriptive words are "phenomenological," "interviews," and "understanding." T40 is closely connected to T2 *identity* and T58

race: African American. Thus, the rise of these topics suggests a shift in the research efforts of PhD students toward qualitative methods and issues of identity and race.

Another popular topic is *adolescence* (T48). We hypothesize that this topic's trend is likely connected to the launch of Add Health in the mid-1990s. The National Longitudinal Study of Adolescent to Adult Health is a representative longitudinal study tracing the transition of then-students to young adults and including social networks as well as various other contextual data, which suited the needs of the scientific community and resulted in a large amount of dissertations and publications (Add Health 2018). T32 *work*—described by concepts like "work," "employee," "job_satisfaction," "organizational_commitment," and "transformational_leadership"—also saw a stark rise but was limited to a relatively

small period, from 2007 to 2012. This trend may be related to the financial crisis and insecure employment opportunities for large parts of U.S. society.

T25 *theory* and topics related to survey methods are on a generally declining path in the past few decades. In the early 1980s, T25 was by far the most common topic, yet it has declined ever since. In 1980, almost 7 percent of words in a thesis were assigned, on average, to T25 *theory*, but this number was cut in half in the mid-1990s, only to decline further to just above a 2 percent average in 2014. Its most exclusive words include “philosophical,” “moral,” “Habermas,” “rationality,” “theorist,” and “Weber.” However, the topic’s most likely terms are “social” and “concept,” which confound theory and general sociological framing terms. This suggests the decline in theory captured by our model may be conservatively estimated. One potential explanation for this decline may be that PhD students face higher demands to foreground empirical results at the expense of theoretical interpretation, as some contemporary critiques suggest (Besbris and Khan 2017; Swedberg 2017).

Students also wrote less about survey methods (T14 and T36). This may be surprising, given the long history and abundant utilization of surveys in social sciences—arguably one of the largest methodological contributions of social sciences in the past century (Platt 1998). However, (household) surveys have faced reduced cooperation of respondents and higher nonresponse, which has led to declining accuracy and an increased need for imputation and inflation of measurement error (Meyer, Mok, and Sullivan 2015). It stands to reason these developments would affect graduate research, as they have affected social research more generally. However, other quantitative topics, like T50 *modeling*, declined much slower than T14 *survey/scales* and T36 *survey/opinion* in the 2000s, and have regained some popularity in recent years.

The other topics dwindling in students’ favor are mostly located at the periphery of the sociological field and connected to other

disciplines like psychology (e.g., T46 *stress* or T23 *status/groups*). In general, these extra-jurisdictional themes occupied less salient positions from the beginning of our observation period; hence, their decline took place at a lower magnitude than the other trends we discuss.

In summary, between 1980 and 2015, students expressed a growing interest in topics related to the cultural turn (T2 *identity*, T40 *qualitative methods*) and race (T58 *race: African American*). After the heated theoretical debates of the 1970s and early 1980s (Alexander 1987), it seems theory, in contrast, has become an auxiliary tool for empirical graduate research and less of a thought style in its own right. At the same time, students’ writing in survey-related methods mirror the challenges these subfields have faced in the past decade.

The Relation between Topic Choice and Becoming an Advisor in Sociology

In this and the following subsection, we test our hypotheses regarding facets of specialization and their effects on the time it takes students to become advisors in sociology. We begin by discussing targeted specialization, then we turn our attention to focus, novelty, and consistency. To test our hypotheses, we present an event-history model that measures the time between thesis and *first* advisorship (see Table 3).¹⁶ In our sample, 2,345 PhD students advised at least one student with a thesis related to sociology or at a sociology department. Thus, only around 3 percent of all PhDs in our sample end up as a primary advisor for doctoral students later in their career. This underlines that our outcome represents a conservative proxy of academic success, that is, advising assumes having a faculty position as well as responsibility over graduate training.

We can confirm our baseline hypothesis (Hypothesis 0) and affirm that the topic structure of a thesis—the particular combination of subjects, methods, and theories—affects career

Table 3. Event-History Models for Time to First Advisorship in Sociology

	(1)	(2)	(3)	(4)	(5)
Elite	2.060*** (.110)	1.887*** (.107)	1.791*** (.105)	1.782*** (.104)	1.720*** (.102)
Female	.882** (.043)	.883* (.045)	.890* (.049)	.888* (.049)	.906 (.050)
White	1.434*** (.081)	1.418*** (.083)	1.431*** (.087)	1.433*** (.087)	1.454*** (.088)
Insider	1.665*** (.084)	1.612*** (.085)	1.434*** (.082)	1.427*** (.082)	1.427*** (.082)
Pub_Cum	1.550*** (.029)	1.275*** (.030)	1.237*** (.031)	1.061 (.036)	.664*** (.053)
Focus		.979 (.025)	.952 (.032)	.972 (.032)	.974 (.036)
Novelty		1.094*** (.027)	1.099*** (.028)	1.096*** (.028)	1.089** (.034)
Consistency		1.388*** (.025)	1.362*** (.026)	2.112*** (.141)	1.752*** (.165)
T2 Identity			1.584*** (.122)	1.574*** (.121)	1.569*** (.120)
T14 Method: Survey/Scales			.666*** (.072)	.662*** (.072)	.670*** (.073)
T50 Method: Modeling			1.352*** (.104)	1.342*** (.103)	1.293*** (.100)
T58 Race: African American			1.356*** (.102)	1.359*** (.103)	1.387*** (.105)
T59 Health			.730*** (.071)	.721*** (.070)	.716*** (.070)
Consistency_Sqr				.865*** (.018)	.868*** (.032)
Consistency × Pub_Cum					1.762*** (.158)
Consistency_Sqr × Pub_Cum					.909*** (.026)
Focus × Pub_Cum					.996 (.023)
Novelty × Pub_Cum					1.008 (.024)
N	82,363	79,137	79,137	79,137	79,137
AIC	27,960	25,417	25,018	24,960	24,801

Note: All models use weights as given by official SED numbers (see Part D of the online supplement), with exponentiated coefficients (hazard ratios) and standard errors in parentheses. Models presented here use only the strictest threshold for targeted specialization (i.e., students who focus more on a topic than 90 percent of their peers). To facilitate readability, we depict the effect of specific topics only if it is significant ($p < .001$) across all thresholds (see Part B of the online supplement for further details on targeted specialization). Note that the reduced number of observations in all models except (1) is due to the construction of the novelty metric (the window comprises three years). Robustness checks for this and other choices are reported in Part F of the online supplement.

* $p < .05$; ** $p < .01$; *** $p < .001$ (two-tailed).

chances. Figure 4 depicts five targeted specializations that are strongly associated with becoming an advisor. Each of these topics is highly significant ($p < .001$) across all thresholds of the targeted specialization variable.¹⁷

Methods matter. Specializing in a specific topic reveals mixed effects. Three topics affect a student’s chances of becoming an advisor positively, and two decrease the likelihood of advisorship. T14 *survey scales* had the

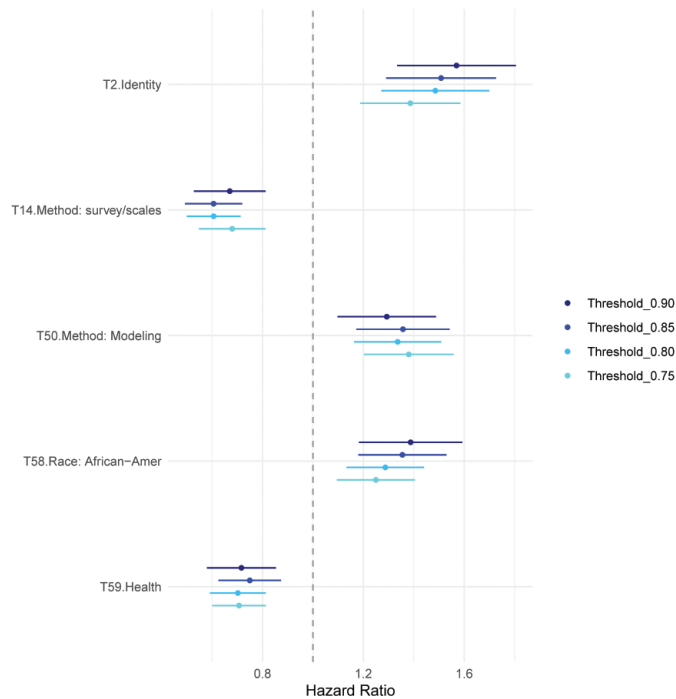


Figure 4. Hazard Ratios for Becoming an Advisor in Sociology

Note: Subset of the event-history model presented in Table 3 (Model 5). Effects of targeted specialization on time it takes students to become advisors for the first time. All topics depicted have a p -value $< .001$ across all thresholds used to construct targeted specialization (see also Section B of the online supplement).

strongest negative effect. T14 was also among the most declining topics, which we linked to severe contemporary challenges with developing surveys (Meyer et al. 2015). However, the negative relationship to advisorship may also stem from the wide range of applications of this technical skill. Future research should explore the extent to which specialization in this topic, and others with similar technical content, may provide differential job opportunities in academia and industry.

In contrast with surveys, T50 *modeling* has a strong positive effect. This topic represents general statistical approaches applied in dissertations and is semantically close to a diverse set of subject matters (see Figure 2). The effects of T14 *survey scales* and T50 *modeling* should be read together as suggesting a possible shift in dominant quantitative research practices and may even represent a turn toward “data science.”

Cultural turn, identity, and race. T2 *identity* is among the most popular and trending topics for sociology PhD students. This topic reflects the cultural turn, one of the major currents in social sciences in the past decades (Alexander 1988; Jacobs and Spillman 2005). Our results reveal that the probability of becoming an advisor rises considerably when a PhD student writes about identity, culture, or discourse. The effect is stronger the more a student focuses on T2.

T58 *race: African American* also significantly increased a student’s chance of becoming an advisor. The effect of T58 also grows with greater focus on it. Note that the specific topic about race our model captures tends to be combined with T2 *identity* (Figure 2). However, the social construction of race as a sociological object of study in U.S. scholarship predates the cultural turn (cf. Du Bois [1897] 2020). Sociological debates about race

in the United States span far more varied intellectual histories than those captured in the cultural turn. These histories are to this day in a state of flux, revision, and recovery (Magubane 2016; Morris 2015).

Neighboring fields. Finally, we find that T59 *health* is negatively associated with the likelihood of becoming an advisor of sociology-related dissertations. Across all thresholds of the targeted specialization covariate, focusing on T59, which has as distinctive terms “HIV,” “obesity,” “health_behavior,” and “immune_deficiency,” reduces chances to advise PhD students by almost 30 percent. Despite its growing popularity in sociology-related dissertations (see Figure 3), T59 is not at the core of the sociological field (see Figure 2). At the same time, social theories seldom inform public health studies (Pound and Campbell 2015). T59 may represent a changing academic environment with more fluid research demands, to which sociology—with its permeable boundaries—is particularly sensitive. Yet, for the observed time period, students specializing in T59 *health* have lower chances of becoming advisors of sociology-related PhDs.¹⁸

To assess the magnitude of the reported effects, Figure 5 compares targeted specialization to the influence of race and gender. Race and gender are two of the most formative dimensions in U.S. society, stratifying social hierarchy in general (McCall 2002), and academe in particular (van den Besselaar and Sandström 2016). Our models confirm that female, minority PhD graduates have a significantly lower likelihood of becoming an advisor, net of an individual’s merit in terms of publications (see Table 3). Thus, racial, and to a lesser degree, gender inequalities leave their footprint in sociology departments, despite the field’s growing diversity (see Figure 1). Comparing these to T50 *modeling*, Figure 5 shows T50’s effect has a similar magnitude.¹⁹ Of course, this does not mean these inequalities can be compensated for through topic choice in a dissertation alone. Instead, this result points to the importance of studying the

interplay of topic choice, status, and inequality in academia (even more for a scholar’s entire record, see Bandelj 2019; Key and Sumner 2019; Koppman and Leahey 2019).

Summing up our results on targeted specialization, we find that topics related to identity, race, and quantitative modeling positively contribute to students’ chances of becoming an advisor in sociology. In contrast, students specializing in survey methods and health in their theses are less likely to advise sociology-related PhDs. This sheds new light on which sociological topics garner influential field positions.

The Influence of Focus, Novelty, and Consistency on Becoming an Advisor in Sociology

In addition to the influence of targeted specialization, we explore three facets of specialization discussed prominently in prior literature on academics’, and more generally, creative workers’, success: focus, novelty, and consistency. By breaking specialization into these analytic dimensions, we are able to identify which facet of specialization is most related to becoming an advisor, and net of the specific topic choices made by PhD students.

Regarding Hypothesis 1, we find no significant effect ($p < .05$) of focus in our models (Table 3). The effect of focus vanishes after considering the influence of targeted specialization and other facets of specialization that are missing controls in previous studies (e.g., Conti et al. 2014; Leahey 2007; Schilling and Green 2011; Teodoridis et al. 2019).²⁰

Hypothesis 2 asks whether novel or traditional combinations of topics in a thesis promote PhD graduates’ career chances. Across all models, students integrating novel combinations of topics in their thesis have close to a 10 percent higher chance of advising sociology PhDs themselves. Again, this is net of other forms of specialization. The sociological field thus seems to reward unusual blends of thought styles expressed in early stages of a scholar’s career. Of course, the specific chains of mechanisms that connect early-stage

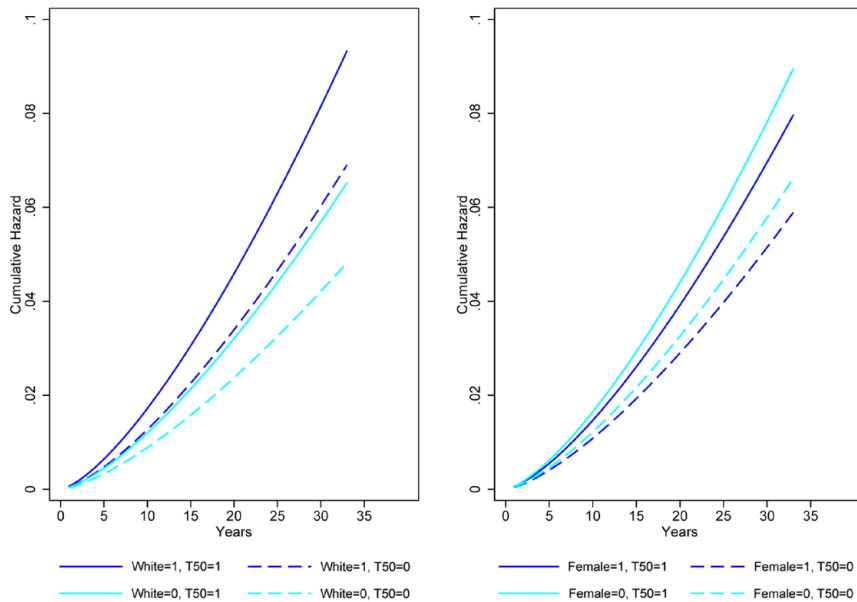


Figure 5. Cumulative Hazards for Time to First Advisorship, by Gender, Race, and Targeted Specialization in *T50 Methods: Modeling*

innovation and later-stage career success will require further study. However, our results are in keeping with Leahey and Moody's (2014) findings that sociology rewards innovations, such as subfield integration.

Our final hypothesis (Hypothesis 3) concerns consistency, and switches the text source from the dissertation to a student's publication track. We find scholars who publish articles consistent with their dissertation have a better chance of becoming advisors. The size of this effect is comparable to the largest effects of specific topic choices. However, the effect is particularly strong for moderate values of consistency and marginally decreases for higher values (i.e., has a negative squared effect). This supports prior work showing that a consistent identity is essential to promote novel work (Trapido 2015) or unusual methods (Koppman and Leahey 2019). Our finding adds further nuance, however, by suggesting there are diminishing returns to identity consistency.

The consistency metric rests on students' number of publications during their careers.

Therefore, we investigate the relationship between publication output over students' careers and their identity consistency.²¹ Figure 6 demonstrates that high productivity can *decrease* the likelihood of mentorship at low levels of consistency; in other words, a dramatically changing research identity is not rewarded. In contrast, only a small number of publications with a consistent thematic angle (in reference to the thesis) provides better chances to achieve a position as an advisor. The relation between identity consistency and productivity, therefore, can describe significant differences in career outcomes.

DISCUSSION

Even the most experienced scientists perceive their discipline from a necessarily limited individual perspective. The sociology of science (Merton 1973), and, more recently, science of science (Fortunato et al. 2018; Wang and Barabási 2021), provide a more comprehensive bird's eye view on academe's inner-workings. A key concern in these areas of

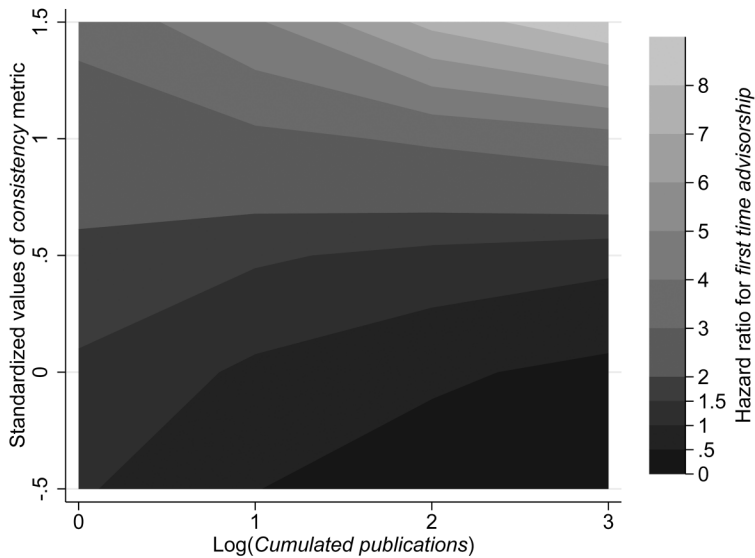


Figure 6. Marginal Effects of Interaction between Consistency and Number of Publications
Note: Effects taken from Model 5 (Table 3). Shades of gray depict hazard ratios for becoming an advisor.

study is the fierce selection process inherent to all scientific fields, that is, who achieves an academic career and by what strategy? In this article, we disentangle distinct facets of specialization as potential career strategies (targeted topics, focus, novelty, and consistency), provide a consistent methodological frame to measure them (structural topic modeling), and use an understudied population (PhD students) to capture how specialization strategies used at an early career stage relate to long-term success in academia.

Our results show across a number of robustness checks that it is not topical focus of a thesis that provides career advantages, but targeted specialization and novel combinations. In particular, theses closely related to the cultural turn (T2 *identity*), statistical approaches (T50 *modeling*), and race (T58 *race: African American*) correspond with rare ascents to mentorship. Dissertations targeting *survey methods* (T14) are, in contrast, less likely to bring the graduate an advisorship. In addition, we find that departures from tradition and embracing novel combinations of topics help graduates become advisors. Last, and most

important, we find that students whose ensuing publications echo and elaborate their dissertations exhibit a consistent research identity and have a considerably higher chance of becoming an advisor. Whereas there are diminishing returns to consistency for highly productive scholars, a relatively small number of consistent publications significantly improves the likelihood of achieving an advisor position, adding important nuance to the publish or perish imperative.

The social sciences are not a favored source of case studies in the sociology of science (Leahey 2008). Yet, we chose to study sociology because of the long-standing polemic in this field regarding specialization. On the one hand, specialization has been recognized as a necessary strategy in a massified field with narrowing job opportunities (Collins 1986). On the other hand, specialization is said to be responsible for the discipline's fragmentation (Coser 1990; Gitlin 1990). In our field, specialization is a multifaceted, contentious, high-stakes career strategy. Our approach was to analytically and empirically distinguish these facets of

specialization. With the exception of targeted specialization, the facets are abstract enough to provide a common analytic framework that can be applied to multiple academic fields.

There are many compelling prospects for such a comparative study of academic specialization. The first one is that facets of specialization may vary in relevance across fields that participate in similar intellectual movements. For instance, academic shifts like the cultural turn (Jacobs and Spillman 2005) encompass the social sciences as a whole, with fits and starts specific to each discipline. Yet, we do not know if underneath this shared thematic surface lie different relations between focus, innovation, consistency, and academic success. Does the combination of novel and consistent specialization yield similar advantages for career prospects in other disciplines? And what about specific topic choices? We have yet to uncover how common themes come to express distinct ways of vying for recognition across academic fields.

A second prospect of a comparative sociology of specialization is the study of migration. Our interest in the reproduction of the sociology field led us to narrowly focus on who becomes an advisor in a sociology department or of a sociology-related dissertation. However, one could imagine other outcomes of interest, such as becoming an advisor in a neighboring or distal field, or becoming an advisor in a focused or eclectic set of topics. On this point, there is not only an opportunity to better understand academics' individual projects, but also the way the academic system distributes rewards and risks in the form of migration patterns.

A third opportunity of this prospective research program would be the comparative study of field relations to government and industry. In a prior study of education research, we found that graduate students grew more interested in interpretive research approaches just as pressure mounted from the government to norm the field according to the standard of randomized control trials (Munoz-Najar Galvez et al. 2020). These results were suggestive of the emergence of

a particular configuration of field autonomy, expressed in how education graduate students chose to specialize. Fields with different relations to industry and government may achieve autonomy in varied ways, which, in turn, should lead students to seek different forms of specialization.

In focusing on a long-term career outcome of specialization (i.e., advisorship), we omitted the lived conditions of scholars that may lead them to choose and sustain this strategy. Future research will do well to embed specialization in a broader system of career checkpoints—spanning from undergraduate to tenure (Meyers et al. 2018)—working arrangements (Cech and Blair-Loy 2014), and family formations (Wolfinger, Mason, and Goulden 2009). In addition, linking PhD students' initial thematic tracks with alternative outcomes could prove fruitful. People who have gained a sure footing in academia as an advisor might also occupy other important positions, for example, in councils, committees, or media discourse. Thus, connecting the dots of scientists' professional life course before and after the PhD seems a worthwhile step over and above our findings.

A key limitation of our study is that we only have access to primary advisors to construct our outcome variable.²² In addition, we have no information on student drop-out rates, so we cannot control for selection biases happening *before* graduation. Nevertheless, we believe the focus on graduating students and their timing to becoming a (primary) advisor is a conservative proxy for academic success, given the resources necessary to achieve doctoral mentorship. We also note that our results are limited to the time period and set of students we study. The stability of our results over 30 years suggests much of this pattern may still hold, such as the extant divisions in the field and specific topics and strategies that secure advising roles. However, we also acknowledge that the period from 2015 to 2021 has been especially tumultuous socially and has likely resulted in further change. It may be that the extant political context has resulted in a sea change for hiring and what matters.

Acknowledgments

Special thanks to David Jurgens, Bas Hofstra, and Vamshi Krishna for help with data collection and architecture. Our gratitude as well to Joshua Gagne, Charles Gomez, Patrick Kaminski, E. Nadia Kline, Benjamin Keep, Elina Mäkinen, Richard Münch, Peter McMahan, and Michael Windzio. These data were collected by the Mimir Project, conducted at Stanford University.

Funding

The data collection on this project was partly funded by the Office of the President at Stanford University and other aspects were partly funded by the National Science Foundation [Awards 1827477, SMA1829240, and 1633036]. We are also grateful for funding received from the Fulbright Program, the Fritz Thyssen Foundation, and the German Federal Ministry of Education and Research [Award 01PU17021C].

ORCID iDs

Raphael H. Heiberger  <https://orcid.org/0000-0003-3465-7214>
 Sebastian Munoz-Najar Galvez  <https://orcid.org/0000-0002-4402-9564>
 Daniel A. McFarland  <https://orcid.org/0000-0002-6805-0798>

Notes

1. To be sure, writing a thesis reflects earlier educational choices. College major selection, for example, has profound effects on educational pathways (Morgan, Gelbgiser, and Weeden 2013).
2. We present our hypotheses on specialization with a dichotomy of hypothesis and antithesis, because extant literature does not provide enough backing to any one claim. Our intention is to be transparent and acknowledge the exploratory nature of our hypotheses.
3. Code and data to construct our metrics and replicate our findings can be found on GitHub (<https://github.com/RapHei/Facets-of-specialization>).
4. We derived the departments of the dissertations listed in ProQuest through the following procedure. On a first iteration, we changed each department name to one of the National Research Council's (NRC) listed department labels. However, the department names in ProQuest are entered manually, so there are plenty of spelling mistakes. Consequently, we used fuzzy string matching in a second iteration to find all cases where the ProQuest department is 90 percent similar to the name in the NRC department, and we automatically treat those names as the same. A manual analysis showed this is 100 percent accurate, because most differences above 90 percent are due to spelling or uncommon

suffices. For frequent names in the 70 to 89 percent range, we manually verified each canonicalization from ProQuest to NRC, rejecting those that were invalid. All dissertations whose department name could not be mapped to an NRC department label had their department inferred as if it had not been listed.

5. The sociology category makes up 50.5 percent of all theses, followed by psychology with 13.5 percent, education (12.7 percent), and the broader category of social and behavioral sciences (12 percent).
6. Thus, STM allow us to consider for each word a document-specific distribution based on its covariates, not only on a general distribution that is the same for all documents, as in the most common variety of topic models, the Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003). The incorporation of covariates improves topic quality substantially (Roberts et al. 2016).
7. To identify advisors' departments, we first calculate the frequency of each student's department per advisor. We then pick the department with the highest frequency as the advisor's disciplinary background. If frequencies are tied, we pick the department at which advisors first mentored, that is, the department of their earliest activity as advisor.
8. It is therefore an even stricter metric for academic success than, for instance, being a professor, because advising signifies many resources (time, reputation, open positions) that not all faculty members at all universities possess. Most PhDs do not go on to become professors at PhD-granting universities but continue their careers at high schools, community colleges, think tanks, policy groups, colleges, or in industry. The relatively low number of 3 percent of students who become advisors reflects these circumstances. In addition, note that individuals with a sociology-related thesis advising students with no reference to sociology would also not appear as advisors in our data, because they do not contribute to sociology's reproduction.
9. The matching process utilizes the metadata of ProQuest and Web of Science (WoS). We use the disambiguated author clusters of Levin and colleagues (2012), which identify groups of publications in WoS authored by the same persons. For matching names semantically, our method is robust to minor typographic errors (as information is entered manually in ProQuest) and enables us to recognize variants of names (e.g., David or Dave) and abbreviations of individuals' first and middle names. The process ensures that one author from either side is not linked to more than one author from the other dataset. As with every complex data operation, the result may be noise. Therefore, we take additional steps to reduce potential mismatches: (1) We use the graduation data given in ProQuest and restrict WoS matches to individuals with a fitting publication history, that is, we exclude individuals whose

- nearest publication is 15 years after or 10 years before graduating. (2) We also refrain from matching individuals with most publications before their graduation, except in cases with evidence that those publications happened in collaboration with their advisor. (3) We consider not only names' similarity, but also textual similarity (cosine) between a student's thesis and publications as well as institutional information, that is, the number of articles on which the university of the WoS author is one of the same institution(s) from the ProQuest database (as an advisor or advisee).
10. We also had recourse to NRC sociology department rankings in 2010, 1993, and 1982. Those are very similar to *U.S. News and World Report* rankings, but they vary in measurement from decade to decade and have significant gaps in department coverage. Thus, we rely on the *U.S. News* rankings from 1992 to 2015 as they afford the best means of identifying elite versus non-elite departments over time. Because we lack rankings prior to 1992, our models assume 1980 rankings hold steady for 12 years. This assumption is not ideal, but the rankings are very stable. *U.S. News* rankings for sociology are correlated at .98 from year to year, and when we consider rankings 12 years apart, they are correlated, on average, at .86 (Spearman rank correlation).
 11. We matched data from the U.S. Census and U.S. Social Security Administration to a dataset containing names, race, and gender of scholars from a specific university (Rawlings et al. 2015). This allows us to train a threshold algorithm to estimate race and gender based on names and to classify advisees in the ProQuest data using these thresholds. Gender is assigned as male and female, and race is classified into White, Asian, Hispanic, and other race. The "other race" category is a residual category and combines African Americans, Native Americans, and other racial categories not captured by the first three, before we binarize it into "White" and "minorities" for our analysis. The algorithm yields high correlations between self-reported and predicted gender and race. We matched 91.3 percent of advisees to a gender (i.e., 8.7 percent with an "unknown" gender) with 95.6 percent accuracy; we matched 89 percent of advisees to a race with 91.2 percent accuracy. The number of "unknown" is given for the specific sample used in this article, but the accuracy is given for all ProQuest data on which we applied the procedure.
 12. As is often the case for event data, our data are right-censored, that is, we do not know if a student in our sample advised a PhD thesis after 2015. One of the major advantages of EHA is that it treats censored observations differently from uncensored observations and so "avoids the problems associated with traditional regression-based techniques" when it comes to the issue of censoring (Box-Steffensmeier and Jones 1997:1430).
 13. Many of the topics are remarkably close to classification schemes used by other researchers. Topics we identify in sociology-related theses (e.g., "religion," "demography," "family," "work," "knowledge") mirror specialty areas used by Moody (2004) as well as Cambridge Scientific Abstracts classifications used by Leahey (2007). Other topics resemble sections of the American Sociological Association (e.g., "crime," "economic sociology," "law"). Some distinctions stand out, such as T58 *race: African American* being a topic but not a subject code; or *social networks* (T41) being distinct from *organizations* (T27).
 14. Trends for all topics are depicted in Part E of the online supplement.
 15. This is an identity topic, which we interpret in relation to the period when it rises in popularity, as well as in relation to its content words. Identity was not discovered in sociology with the cultural turn. This topic does not encompass all sociological debates about identity.
 16. Many advisors accompany more than one student to PhD honors. Around 1,600 advisors in our sample had more than one advisee. To account for the occurrence of multiple advisorships, we use repeated event models (conditional gap time models, as elaborated in Part D of the online supplement). We report the results together with other robustness checks in Part F of the online supplement.
 17. Part B of the online supplement provides a detailed discussion of varying thresholds. The results are also very similar across a number of additional robustness checks (see Part F of the online supplement).
 18. T59 *health* might also point to an unmeasured but relevant mechanism. We cannot control if students focusing on cross-disciplinary topics like T59 find employment in interdisciplinary research centers, in which they advise no (sociology) students. We are thankful to an anonymous reviewer pointing us to this possibility.
 19. The curves for other topics reported in Figure 4 are similar and can be provided upon request. They can also be created with the code from the paper's repository (<https://github.com/RapHei/Facets-of-specialization>).
 20. To ensure our result is not specific to the construction of topical focus with the Herfindahl index, we used other ways to measure focus, with the same result (see Part F of the online supplement).
 21. The polynomials for focus and novelty showed no significant results. We also tested all possible interactions across our core strategies of focus, novelty, and consistency (tables available upon request). Again, we find no significant effects. Therefore, it is not combined strategies that lead students to advisorship, but multiple, independent routes. These findings are consistent with a heterogeneous

field responsive to changing populations and social issues. It also suggests sociology is a field that rewards exploration and the establishment of new lines of research, rather than rewarding engagement in normal science and furthering extant traditions.

22. In the case of co-chairs, that is, two equally ranking advisors, ProQuest provides only the one who is entered first in the system. This list order was decided by the student. To ensure the ordering of primary advisors is not biased systematically, we checked if the ranking is influenced by alphabetic order, but we did not find any such effects (results available upon request).

References

- Abbott, Andrew. 2001. *Chaos of Disciplines*. Chicago: University of Chicago Press.
- Add Health. 2018. "Publications" (<http://www.cpc.unc.edu/projects/addhealth/publications>).
- Alexander, Jeffrey C. 1987. *Sociological Theory since 1945*. London, UK: Routledge.
- Alexander, Jeffrey C. 1988. "The New Theoretical Movement." Pp. 77–101 in *Handbook of Sociology*, edited by N. J. Smelser. Beverly Hills, CA: Sage Publications.
- American Sociological Association (ASA). 2021a. "Mean Number of Tenured and Tenure-Track Sociology Faculty" (<https://www.asanet.org/academic-professional-resources/data-about-disipline/data-dashboard/sociology-programs/mean-number-tenured-and-tenure-track-sociology-faculty-institution-type>).
- American Sociological Association (ASA). 2021b. "Doctorates Awarded in Sociology, by Gender" (<https://www.asanet.org/academic-professional-resources/data-about-disipline/data-dashboard/degrees-awarded/doctorates-awarded-sociology-gender>).
- American Sociological Association (ASA). 2021c. "Doctorates Awarded in Sociology, by Race or Ethnicity" (<https://www.asanet.org/academic-professional-resources/data-about-disipline/data-dashboard/degrees-awarded/doctorates-awarded-sociology-race-or-ethnicity>).
- Amorim, Leila DAF, and Jianwen Cai. 2015. "Modelling Recurrent Events: A Tutorial for Analysis in Epidemiology." *International Journal of Epidemiology* 44(1):324–33.
- Anderson, Ashton, Daniel A. McFarland, and Dan Jurafsky. 2012. "Towards a Computational History of the ACL: 1980–2008." Pp. 13–21 in *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*. Stroudsburg, PA: Association for Computational Linguistics.
- Antons, David, Amol M. Joshi, and Torsten Oliver Salge. 2019. "Content, Contribution, and Knowledge Consumption: Uncovering Hidden Topic Structure and Rhetorical Signals in Scientific Texts." *Journal of Management* 45(7):3035–76.
- Astley, W. Graham. 1985. "Administrative Science as Socially Constructed Truth." *Administrative Science Quarterly* 30(4):497–513.
- Balsmeier, Benjamin, and Maikel Pellens. 2014. "Who Makes, Who Breaks: Which Scientists Stay in Academia?" *Economics Letters* 122(2):229–32.
- Bandelj, Nina. 2019. "Academic Familism, Spillover Prestige and Gender Segregation in Sociology Subfields: The Trajectory of Economic Sociology." *The American Sociologist* 50(4):488–508.
- Besbris, Max, and Shamus Khan. 2017. "Less Theory. More Description." *Sociological Theory* 35(2):147–53.
- Best, Steven. 2007. "Culture Turn." *The Blackwell Encyclopedia of Sociology*. Hoboken, NJ: Wiley-Blackwell.
- Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55(4):77–84.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *The Journal of Machine Learning Research* 3:993–1022.
- Blossfeld, Hans-Peter, Götz Rohwer, and Thorsten Schneider. 2019. *Event History Analysis with Stata*. London, UK: Routledge.
- Bourdieu, Pierre. 1988. *Homo Academicus*. Stanford, CA: Stanford University Press.
- Box-Steffensmeier, Janet M., and Bradford S. Jones. 1997. "Time Is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41(4):1414–61.
- Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. Cambridge, UK: Cambridge University Press.
- Burris, Val. 2004. "The Academic Caste System: Prestige Hierarchies in PhD Exchange Networks." *American Sociological Review* 69(2):239–64.
- Burt, Ronald S. 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110(2):349–99.
- Cech, Erin A., and Mary Blair-Loy. 2014. "Consequences of Flexibility Stigma among Academic Scientists and Engineers." *Work and Occupations* 41(1):86–110.
- Cech, Erin A., Brian Rubineau, Susan Silbey, and Carroll Seron. 2011. "Professional Role Confidence and Gendered Persistence in Engineering." *American Sociological Review* 76(5):641–66.
- Coleman, James S. 1981. *Longitudinal Data Analysis*. New York: Basic Books.
- Collins, Randall. 1986. "Is 1980s Sociology in the Doll-drums?" *American Journal of Sociology* 91(6):1336–55.
- Collins, Randall. 2002. *The Sociology of Philosophies: A Global Theory of Intellectual Change*, rev. ed. Cambridge, MA: Belknap Press of Harvard University Press.
- Conti, Raffaele, Alfonso Gambardella, and Myriam Mariani. 2014. "Learning to Be Edison: Inventors, Organizations, and Breakthrough Inventions." *Organization Science* 25(3):833–49.

- Correll, Shelley J., Cecilia L. Ridgeway, Ezra W. Zuckerman, Sharon Jank, Sara Jordan-Bloch, and Sandra Nakagawa. 2017. "It's the Conventional Thought That Counts: How Third-Order Inference Produces Status Advantage." *American Sociological Review* 82(2):297–327.
- Coser, Lewis A. 1990. "The Virtues of Dissent in Sociology." Pp. 207–13 in *Sociology in America*. London, UK: Sage Publications.
- Crane, Diana, and Henry Small. 1992. "American Sociology since the Seventies: The Emerging Identity Crisis in the Discipline." Pp. 197–234 in *Sociology and Its Publics: The Forms and Fates of Disciplinary Organization*, edited by T. C. Halliday and M. Janowitz. Chicago: University of Chicago Press.
- Du Bois, W. E. B. [1897] 2020. *The Conservation of Races*. New York: Good Press.
- Evans, James A., and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42(1):21–50.
- Ferguson, John-Paul, and Sharique Hasan. 2013. "Specialization and Career Dynamics: Evidence from the Indian Administrative Service." *Administrative Science Quarterly* 58(2):233–56.
- Fleck, Ludwik. 1979. *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.
- Fleming, Lee. 2001. "Recombinant Uncertainty in Technological Search." *Management Science* 47(1):117–32.
- Fleming, Lee, Santiago Mingo, and David Chen. 2007. "Collaborative Brokerage, Generative Creativity, and Creative Success." *Administrative Science Quarterly* 52(3):443–75.
- Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang, and Albert-László Barabási. 2018. "Science of Science." *Science* 359(6379):eaao0185.
- Foster, Jacob G., Andrey Rzhetsky, and James A. Evans. 2015. "Tradition and Innovation in Scientists' Research Strategies." *American Sociological Review* 80(5):875–908.
- Frickel, Scott, and Neil Gross. 2005. "A General Theory of Scientific/Intellectual Movements." *American Sociological Review* 70(2):204–32.
- Giddens, Anthony. 1976. "Classical Social Theory and the Origins of Modern Sociology." *American Journal of Sociology* 81(4):703–29.
- Gieryn, Thomas F. 1978. "Problem Retention and Problem Change in Science." *Sociological Inquiry* 48(3–4):96–115.
- Gitlin, Todd. 1990. "Sociology for Whom? Criticism for Whom?" Pp. 214–26 in *Sociology in America*. London, UK: SAGE Publications.
- Grant, Linda, Kathryn B. Ward, and Xue Lan Rong. 1987. "Is There an Association between Gender and Methods in Sociological Research?" *American Sociological Review* 52(6):856–62.
- Griffiths, Thomas L., and Mark Steyvers. 2004. "Finding Scientific Topics." *Proceedings of the National Academy of Sciences* 101:5228–35.
- Hand, Carl M., and Ben Judkins. 1999. "Disciplinary Schisms: Subspecialty 'Drift' and the Fragmentation of Sociology." *The American Sociologist* 30(1):18–36.
- Hannan, Michael T., Gaël Le Mens, Greta Hsu, Balázs Kovács, Giacomo Negro, László Pólos, Elizabeth Pontikes, and Amanda J. Sharkey. 2019. *Concepts and Categories: Foundations for Sociological and Cultural Analysis*. New York: Columbia University Press.
- Hofstra, Bas, Rense Corten, Frank van Tubergen, and Nicole B. Ellison. 2017. "Sources of Segregation in Social Networks: A Novel Approach Using Facebook." *American Sociological Review* 82(3):625–56.
- Hofstra, Bas, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland. 2020. "The Diversity–Innovation Paradox in Science." *Proceedings of the National Academy of Sciences* 117(17):9284–91.
- Hofstra, Bas, and Nick C. de Schipper. 2018. "Predicting Ethnicity with First Names in Online Social Media Networks." *Big Data & Society* (<https://doi.org/10.1177/2053951718761141>).
- Holmwood, John. 2010. "Sociology's Misfortune: Disciplines, Interdisciplinarity and the Impact of Audit Culture." *The British Journal of Sociology* 61(4):639–58.
- Hoppe, Travis A., Aviva Litovitz, Kristine A. Willis, Rebecca A. Meseroll, Matthew J. Perkins, B. Ian Hutchins, Alison F. Davis, Michael S. Lauer, Hannah A. Valentine, James M. Anderson, and George M. Santangelo. 2019. "Topic Choice Contributes to the Lower Rate of NIH Awards to African-American/Black Scientists." *Science Advances* 5(10) (<https://doi.org/10.1126/sciadv.aaw7238>).
- Huang, Junming, Alexander J. Gates, Roberta Sinatra, and Albert-László Barabási. 2020. "Historical Comparison of Gender Inequality in Scientific Careers across Countries and Disciplines." *Proceedings of the National Academy of Sciences* 117(9):4609–16.
- Inglis, Matthew, and Colin Foster. 2018. "Five Decades of Mathematics Education Research." *Journal for Research in Mathematics Education* 49(4):462–500.
- Jacobs, Mark D., and Lyn Spillman. 2005. "Cultural Sociology at the Crossroads of the Discipline." *Poetics* 33(1):1–14.
- Jacomy, Mathieu, Tommaso Venturini, Sebastian Heymann, and Mathieu Bastian. 2014. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLoS ONE* 9(6):e98679.
- Key, Ellen M., and Jane Lawrence Sumner. 2019. "You Research Like a Girl: Gendered Research Agendas and Their Implications." *PS: Political Science & Politics* 52(4):663–68.
- Knorr Cetina, Karin. 2009. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

- Koppman, Sharon, and Erin E. Leahey. 2019. "Who Moves to the Methodological Edge? Factors That Encourage Scientists to Use Unconventional Methods." *Research Policy* 48(9):103807.
- Kuhn, Thomas S. 1959. "The Essential Tension: Tradition and Innovation in Scientific Research." Pp. 162–74 in *The Third University of Utah Research Conference on the Identification of Scientific Talent*, edited by C. W. Taylor. Salt Lake City: University of Utah Press.
- Kuhn, Thomas S. 1962. *Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lamont, Michèle, and Virág Molnár. 2002. "The Study of Boundaries in the Social Sciences." *Annual Review of Sociology* 28:167–95.
- Leahey, Erin. 2006. "Transmitting Tricks of the Trade: Advisors and the Development of Research Knowledge." *Teaching Sociology* 34(2):93–110.
- Leahey, Erin. 2007. "Not by Productivity Alone: How Visibility and Specialization Contribute to Academic Earnings." *American Sociological Review* 72(4):533–61.
- Leahey, Erin. 2008. "Methodological Memes and Mores: Toward a Sociology of Social Research." *Annual Review of Sociology* 34(1):33–53.
- Leahey, Erin, Christine M. Beckman, and Taryn L. Stanko. 2017. "Prominent but Less Productive: The Impact of Interdisciplinarity on Scientists' Research." *Administrative Science Quarterly* 62(1):105–39.
- Leahey, Erin, Bruce Keith, and Jason Crockett. 2010. "Specialization and Promotion in an Academic Discipline." *Research in Social Stratification and Mobility* 28(2):135–55.
- Leahey, Erin, and James Moody. 2014. "Sociological Innovation through Subfield Integration." *Social Currents* 1(3):228–56.
- Levin, Michael, Stefan Krawczyk, Steven Bethard, and Dan Jurafsky. 2012. "Citation-Based Bootstrapping for Large-Scale Author Disambiguation." *Journal of the American Society for Information Science and Technology* 63(5):1030–47.
- Lutter, Mark, and Martin Schröder. 2016. "Who Becomes a Tenured Professor, and Why? Panel Data Evidence from German Sociology, 1980–2013." *Research Policy* 45(5):999–1013.
- Magubane, Zine. 2016. "American Sociology's Racial Ontology: Remembering Slavery, Deconstructing Modernity, and Charting the Future of Global Historical Sociology." *Cultural Sociology* 10(3):369–84.
- Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press.
- McCall, Leslie. 2002. *Complex Inequality: Gender, Class and Race in the New Economy*. London, UK: Routledge.
- Merton, Robert K. 1973. *The Sociology of Science*. Chicago: The University of Chicago Press.
- Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29(4):199–226.
- Meyers, Lindsay, Abigail Brown, Liane Moneta-Koehler, and Roger Chalkley. 2018. "Survey of Checkpoints along the Pathway to Diverse Biomedical Research Faculty." *PLoS One* 13(1):e0190606.
- Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45(1):27–45.
- Moody, James. 2004. "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999." *American Sociological Review* 69(2):213–38.
- Moore, Joan. 1990. "Criticism for Whom? Social Criticism and Sociological Elitism." Pp. 227–30 in *Sociology in America*. London, UK: Sage Publications.
- Morgan, Stephen L., Dafna Gelbgiser, and Kim A. Weeden. 2013. "Feeding the Pipeline: Gender, Occupational Plans, and College Major Selection." *Social Science Research* 42(4):989–1005.
- Morris, Aldon. 2015. *The Scholar Denied: W. E. B. Du Bois and the Birth of Modern Sociology*. Oakland: University of California Press.
- Munoz-Najar Galvez, Sebastian, Raphael H. Heiberger, and Daniel McFarland. 2020. "Paradigm Wars Revisited: A Cartography of Graduate Research in the Field of Education (1980–2010)." *American Educational Research Journal* 57(2):612–52.
- National Science Foundation. 2020. "Survey of Earned Doctorates" (<https://ncesdata.nsf.gov/home/>).
- Paré, Anthony, Doreen Starke-Meyerring, and Lynn McAlpine. 2011. "Knowledge and Identity Work in the Supervision of Doctoral Student Writing: Shaping Rhetorical Subjects." Pp. 215–36 in *Writing in Knowledge Societies*, edited by D. Starke-Meyerring, A. Paré, N. Artemeva, M. Horne, and L. Yousoubova. WAC Clearinghouse (<https://doi.org/10.37514/PER-B.2011.2379.2.11>).
- Platt, Jennifer. 1998. *A History of Sociological Research Methods in America, 1920–1960*. Cambridge, UK: Cambridge University Press.
- Pound, Pandora, and Rona Campbell. 2015. "Locating and Applying Sociological Theories of Risk-Taking to Develop Public Health Interventions for Adolescents." *Health Sociology Review* 24(1):64–80.
- Rawlings, Craig M., Daniel McFarland, Linus Dahlander, and Dan Wang. 2015. "Streams of Thought: Knowledge Flows and Intellectual Cohesion in a Multidisciplinary Era." *Social Forces* 93(4):1687–722.
- Rhoades, Lawrence J. 1981. *A History of the American Sociological Association, 1905–1980*. Washington, DC: American Sociological Association.
- Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airolti. 2016. "A Model of Text for Experimentation in the Social Sciences." *Journal of the American Statistical Association* 111(515):988–1003.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–82.

- Schilling, Melissa A., and Elad Green. 2011. "Recombinant Search and Breakthrough Idea Generation: An Analysis of High Impact Papers in the Social Sciences." *Research Policy* 40(10):1321–31.
- Schwemmer, Carsten, and Oliver Wieczorek. 2020. "The Methodological Divide of Sociology: Evidence from Two Decades of Journal Publications." *Sociology* 54(1):3–21.
- Shi, Xiaolin, Jure Leskovec, and Daniel A. McFarland. 2010. "Citing for High Impact." Pp. 49–58 in *Joint Conference on Digital Libraries* (JCDL 2010). Brisbane, Australia (<https://doi.org/10.1145/1816123.1816131>).
- Sinatra, Roberta, Dashun Wang, Pierre Deville, Chaoming Song, and Albert-László Barabási. 2016. "Quantifying the Evolution of Individual Scientific Impact." *Science* 354(6312) (<https://doi.org/10.1126/science.aaf5239>).
- Singer, Judith D., and John B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford, UK: Oxford University Press.
- Swedberg, Richard. 2017. "Theorizing in Sociological Research: A New Perspective, a New Departure?" *Annual Review of Sociology* 43:189–206.
- Teodoridis, Florenta, Michaël Bikard, and Keyvan Vakili. 2019. "Creativity at the Knowledge Frontier: The Impact of Specialization in Fast- and Slow-Paced Domains." *Administrative Science Quarterly* 64(4):894–927.
- Teplitskiy, Misha, Daniel Acuna, Aïda Elamrani-Raoult, Konrad Kording, and James Evans. 2018. "The Sociology of Scientific Validity: How Professional Networks Shape Judgement in Peer Review." *Research Policy* 47(9):1825–41.
- Trapido, Denis. 2015. "How Novelty in Knowledge Earns Recognition: The Role of Consistent Identities." *Research Policy* 44(8):1488–500.
- Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342(6157):468–72.
- van den Besselaar, Peter, and Ulf Sandström. 2016. "Gender Differences in Research Performance and Its Impact on Careers: A Longitudinal Case Study." *Scientometrics* 106(1):143–62.
- Wang, Dashun, and Albert-László Barabási. 2021. *The Science of Science*. Cambridge, UK: Cambridge University Press.
- Warren, John Robert. 2019. "How Much Do You Have to Publish to Get a Job in a Top Sociology Department? Or to Get Tenure? Trends over a Generation." *Sociological Science* 6:172–96.
- Weeber, Stan C. 2006. "Elite versus Mass Sociology: An Elaboration on Sociology's Academic Caste System." *The American Sociologist* 37(4):50–67.
- Wolfinger, Nicholas, Mary A. Mason, and Marc Goulden. 2009. "Stay in the Game: Gender, Family Formation and Alternative Trajectories in the Academic Life Course." *Social Forces* 87(3):1591–621.
- Znaniecki, Florian. 1945. "Controversies in Doctrine and Method." *American Journal of Sociology* 50(6):514–21.
- Zuckerman, Ezra W., Tai-Young Kim, Kalinda Ukanwa, and James von Rittmann. 2003. "Robust Identities or Nonentities? Typecasting in the Feature-Film Labor Market." *American Journal of Sociology* 108(5):1018–74.
- Zuckermann, Harriet. 1977. *Scientific Elite: Nobel Laureates in the United States*. New York: Free Press.

Raphael H. Heiberger is a tenure-track professor at the University of Stuttgart and head of a lab on computational social science. He graduated from the Otto-Friedrich University of Bamberg and was a Fulbright Fellow at UC-Berkeley. His research interests revolve around statistical models, scientific careers, political polarization, and financial markets.

Sebastian Munoz-Najar Galvez is the Bluhm Family Assistant Professor of Data Science and Education at the Harvard Graduate School of Education. He studies how research agendas form, and how they accrue public value.

Daniel A. McFarland is a professor of education, and by courtesy, sociology and organizational behavior at Stanford University. His current research is focused on the sociology of science and knowledge innovation.