WWU Honors College Senior Projects

WWU Graduate and Undergraduate Scholarship

Spring 2021

The Secret History: The Statistics and Sociology Behind Fiction

Sam Wick
Western Washington University

Recommended Citation

Introduction:

Fiction is meant to, in some capacity, reflect reality, so it is interesting to assess how well it does just that. This paper seeks to examine a fictional character and use the demographics that they occupy as a template by which to look at data acquired by the National Health and Nutrition Examination Survey (NHANES) from the 2017-2018 survey cycle. This is the aim of this paper due to the integral role that fiction plays in culture and society, both in how it shapes individuals' interactions with one another and the larger world, and how it shapes their expectations and understanding of both.

This paper will be addressing the depiction of a contemporary individual in a contemporary piece of written, realistic fiction. The emphasis on both character and media being contemporary is necessary and relevant, as the referential statistics being used for analysis will be from the 2017-2018 NHANES survey cycle.

The fictional subject for comparison against 2017-18 NHANES data will be the protagonist from Donna Tartt's *The Secret History,* twenty-eight year old Richard Papen. The novel, classified as "dark academia", was first published in 1992 and is set to take place in the 1980s. Papen tells the novel's tale retrospectively, centering on his time in college beginning when he was nineteen. Therefore, for the purposes of this paper, Papen will be regarded as being nineteen, as that is the context in which the most information is given about him.

Papen was chosen due to the fact that early on in the novel, the reader learns that he consistently lied to others about his personal background in order to fit in at the elite university he was attending. He fabricated a new upbringing that made him appear more interesting and as though he came from a more socio-economically well-off background. For instance, he hides the

fact that he is receiving financial aid from the university, partly because a professor with whom he wants to take classes won't accept students into his class who receive financial aid.

This context provides another level of intrigue to Papen's character and has interesting sociological implications. If able to, it would be interesting to compare the statistical outcomes of people from the background that Papen actually has and compare it to the one he claims to have. However, this would be a secondary aim of this paper.

Literature Review:

There is much existing literature on the importance of fiction in general, looking at how it impacts people and their life decisions. For instance, Lindy Orthia's 2019 article "How Does Science Fiction Television Shape Fans' Relationships to Science? Results from a Survey of 575 'Doctor Who' Viewers" does just as the title would suggest, and questions how viewers of science fiction interact with science as a result of their viewership. Orthia finds that many viewers of the television show *Doctor Who* specifically did develop a greater interest in science as a result of watching the show.

While literature on the specific topic this paper will be covering appears limited or nonexistent, examples like that above show not only precedent for the examination to take place, but as will be seen with other incidences, a general consensus of the importance of fiction in a grand cultural and social context. It should be noted, however, that due to different disciplinary or colloquial uses of terms used to search for previous literature — such as the word "fiction" which is often used in article titles in the phrase "Fact or Fiction" — finding relevant previous works is even more challenging.

Those works that do exist and bear some resemblance to this topic, have the caveat of having their own specific focus. For instance, one can find an article specifically looking at the accuracy of representation of women in Shakespeare's plays (Beckwith 2015) or the relationship of twentieth-century Canadian fiction had with modernism (Hill 2012).

Others, like Arthur Jacobs and Roel Willems in their 2012 article "The Fictive Brain: Neurocognitive Correlates of Engagement in Literature", and Anna Abraham, D. Yves von Cramon, and Ricarda Schubotz's 2008 article "Meeting George Bush versus Meeting Cinderella: The Neural Response When Telling Apart What is Real from What is Fictional in the Context of Our Reality" look at how people interact with the notion of fiction, so to say. Interestingly, the human brain is awful at distinguishing between "real" people and those in fiction, causing a sort of blending that places "real" and fictional people in the same category, and for them, as a result, be regarded in the same manner. This lends phenomenon informs the examination and study of how people interact with fiction, at least in part. This, in addition to the tendency of people to take on characteristics and tendencies of those around them both support the importance of fiction in a socio-cultural context, and that importance in turn supports the necessity of its study.

Karen Dill-Shackleford, Cynthia Vinney, and Kristin Hopper-Losenicky also examine this in their 2016 article "Connecting the dots between fantasy and reality: The social psychology of our engagement with fictional narrative and its functional value" which looks specifically at the role of fiction in the day-to-day human experience, including the function that fiction serves. This includes looking at how it can encourage people to question their beliefs and values, and improve social skills, like connecting to others in general, but also developing empathy and acceptance for those different than oneself. This, stemming from the finding that when people consume more diverse media, it has a very similar effect as having a more diverse group of

friends, which connects back to the human brain's inability to distinguish between "real" and fictional persons. They also mention the experience of consuming fiction as a sort of vicarious experience, and how that allows viewers to more thoroughly connect with characters. This further establishes the integral role that fiction plays in the social development of people, and its importance in that regard, and by extension the previously referenced larger socio-cultural context. These pieces point to the importance of portrayal of people in media and suggest that accuracy or inaccuracy in such media will be consumed by viewers, and in some capacity internalized, just as the rest of the media is.

The ability to establish precedent for examining the realistic validity of fiction, in some capacity, does bode well, as it shows that there is already some confidence in the value of doing such. It is apparent that fiction is important, and researchers do care about its accuracy, though particular field and interest will of course play a large part in what aspect or work they examine.

Methods:

Data are being taken from the 2017-2018 survey cycle of the National Health and Nutritional Examination Survey (NHANES). The NHANES program is a subset of the Centers for Disease Control and Prevention (CDC), and performs annual studies of children and adults in the United States. These studies include both interviews with the subjects, as well as performing physical examinations. The goal of NHANES is to produce health statistics to be freely available and used (NHANES).

The character to be used as the entry-point for statistical examination is Richard Papen, the protagonist of Donna Tartt's *A Secret History.* At the time of telling the story, he is twenty-

eight, describing his time at a university in New England, when he was nineteen. He came from

an unassuming background, but fabricated a much more interesting background for himself in

order to appear more interesting and better fit in at the prestigious university he attended.

For the purposes of this paper, Richard Papen will be regarded as being a nineteen year

old, caucasian male, only child from a three-person middle to lower-income household. During

the novel, he has graduated high school and attended college for two years before transferring to

the university he is in for the duration of the novel. It is unknown if he achieved his AA before

transferring, so it will be assumed that he did.

To begin, frequencies will be run in the program SPSS on relevant demographic statistics,

including age, race, gender, household income, number of household members, and highest level

of education achieved. From there, bivariate crosstabulations will be completed gender and

highest level of education, as well as for race and highest level of education. Then, a multivariate

regression will be performed to examine the impact of race and income on highest level of

education achievement.

From there, it is hoped that more variables will be crosstabulated, such those from the

depression screening questions, as well as potential felon-status information. This, because of the

feelings of isolation Papen experiences in regards to his peers, stemming from being in a class of

less than ten students, including himself and his professor, and the fact that his background is a

fabrication, and his true background is extremely different than that of those around him and he

fears his lies being discovered. On top of this, in the novel he was part of a cooperative murder

of one of those classmates, something revealed in the prologue of the novel, which would likely

have an impact on his mental health, even more so if he had been caught and became a felon as a result of that.

<u>Findings:</u>

Just looking at the key variables, 49.2% of respondents were male, 16.8% of respondents had a family size of three people, and 34% of respondents were non-hispanic white. 1.5% of respondents were 19 years old. 7.5% were between 18 and 24 years old (that being an estimation of the typical age range of persons in college).

6.5% of respondents between the ages of 6 and 19 years old had graduated high school, and that was their highest reported level of education. 3.1% of respondents in that age range had achieved more than high school as their highest level of education. For adults, 20 years old or older, 23.8% had high school graduate, GED or equivalent as their highest level of education. 31.9% reported some college or an AA degree as their highest level of education. 24% reported college graduate or above as their highest level of education.

In a crosstabulation examining gender and the highest level of education achieved by people 20 years of age and older, a result emerged that didn't appear when looking at the the highest level of education of people aged 6 to 19 years old.

In the older group, there was a result of statistical significance with 99% confidence showing that respondents who were female reported a higher level of education. Male respondents had higher percentages in having graduated high school or received a GED or equivalent or lower as their highest level of education. For having any college, an AA, graduated college or above, female respondents had higher percentages. 21.4% of male respondents 20

years or older didn't graduate high school, versus 18.7% of female respondents. Similarly, 53.5%

of male respondents completed some college or above, versus 58.2% of female respondents.

There also appeared to be a 99% confident statistically significant relationship between

the highest level of education achieved by respondents 20 years old and above and race and

Hispanic origin. 46.1% of Mexican American respondents reported that they did not graduate

high school, receive a GED or equivalent as their highest level of education achieved. This,

compared to 37.9% of respondents identified as other Hispanic, 12.6% of non-Hispanic White

respondents, 14.8% of non-Hispanic Black respondents, and 13.6% of respondents identified as

part of other racial groups, including multiracial respondents.

Looking at a multivariate regression, looking to examine potential impact that race and

Hispanic origin as well as annual family income might have on the highest level of education

reported by respondents age 20 and older, there were noteworthy results. As suggested by

previous crosstabulations, it can be said with 99% confidence that race and Hispanic origin have

a statistically significant impact on the highest level of education reported by respondents aged

20 and older. On the other hand, annual family income may have statistical significance, though

it was not quite to the point where one could be 95% confident, so it's much less certain.

Now, ideally I would have been able to crosstabulate some of these things with the

depression screening questions and felon status. NHANES, unfortunately, does not include

questions about felon status or anything similar that could be used as a reasonable substitute. Just

to be sure, I checked other survey cycles, and they have never asked such questions. I was very

much looking forward to that.

Similarly, I had planned to crosstabulate with the depression screening questions, which are, unhelpfully, in a completely isolated dataset. Now, while that could still have proved interesting, after running a frequency on all variables in that set, the numbers in that dataset are so polarized that I don't think I would learn much from using them, especially considering I would have to figure out how to port them into the demographics dataset I pulled everything else from. In every question in the depression screening set, at least 50% of respondents — though typically more, most are about 75-85% — responded the equivalent of "not at all".

My hypothesis going in was that I would see a lot more variability in the depression screening numbers, and that were I able to run felon status, that would return with disproportionately high numbers on the depression screening questions.

There is still potentially a point to be made about such high numbers of respondents answering so similarly, but it would be incredibly speculative. My questioning would be about their honesty in responding — addressing the social stigma — and potentially interrogating the way the questions are phrased. However, that seems very detached from everything else I've done, so there's the question of tying it together, as well.

This would have been a key point of my research, as it would allow me to not only discuss the mental health — specifically depression — as it is potentially related to highest level of education, but also how it relates to being a felon.

Discussion:

While my data didn't lead to as many conclusions I'd hoped to draw, I was able to fulfill the statistical dream: finding statistical significance with 99% confidence. Although it wasn't in

regards to anything surprising, a significant part of any scientific discipline is looking at data to confirm — or question — what people already know or believe they know.

When looking to what I was able to confirm, there does seem to be a strong relationship between racial background and highest level of achievement. That seemed like a given, but the specifics of it drew my attention. Seeing that Mexican Americans were the least likely to get a college degree was not something I had expected to find, in large part due to how often racial lines are draw between white and black, which can forget to include other racial groups.

Further, in not finding what I was looking for in terms of health-outcomes because of the oddly disproportionate skew of the depression screening questions and absence of other questions of interest, my research helps to illustrate those gaps and shortcomings. I have, in some small way, highlighted gaps in knowledge, resulting from oversights in the NHANES survey questions. Among these, notably, is the phrasing of the depression screening questions. Being so direct, while effective on a surface-level, fails to take into account the social element of the questions, and the stigma that surrounds depression and mental health in general.

On that note, that could be an interesting research project in itself, examining how tendencies in those responses have changed over the years. This could lead to interesting results when cross-examined with large social events and upheavals, but also — when newer data is released especially — the contemporary efforts toward de-stigmatizing mental health and raising awareness about it.

Tangent aside, while it's frustrating in the position of a researcher to not find what one set out looking for, it's often unavoidable, especially with limited time and resources. Nonetheless, I have still managed to contribute, maybe not to literature about health, as had been the intention,

but sociology and inequity. The racial disparity in educational achievement has not disappeared, my analysis of the data shows that.

My analysis may not synthesize how the racial disparity has changed and developed, but it does draw attention to the Mexican American population. It also draws attention to the more surprising fact that annual family income had a much less significant impact on highest level of education achieved, versus racial background. This, I found particularly interesting because so often in sociology race and income are linked. While this is often for good cause, this suggests that they still need to be examined separately as well.

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Education level – Adults 20+ * Gender	5569	60.2%	3685	39.8%	9254	100.0%

Education level – Adults 20+ * Gender Crosstabulation

			Gender		Total
			1	2	
Education level – Adults 20+	1	Count	240	239	479
		% within Gender	8.9%	8.3%	8.6%
	2	Count	339	299	638
		% within Gender	12.5%	10.4%	11.5%
	3	Count	671	654	1325
		% within Gender	24.8%	22.8%	23.8%
	4	Count	807	971	1778
		% within Gender	29.9%	33.9%	31.9%
	5	Count	639	697	1336
		% within Gender	23.6%	24.3%	24.0%
	7	Count	0	2	2
		% within Gender	0.0%	0.1%	0.0%
	9	Count	6	5	11
		% within Gender	0.2%	0.2%	0.2%
Total		Count	2702	2867	5569
		% within Gender	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	17.591[a]	6	.007
Likelihood Ratio	18.371	6	.005
Linear-by-Linear Association	6.577	1	.010
N of Valid Cases	5569		

a. 2 cells (14.3%) have expected count less than 5. The minimum expected count is .97.

<div align="center">References</div>

Abraham, Anna, D. Yves von Cramon, and Ricarda I. Schubotz. 2008. "Meeting George Bush

versus Meeting Cinderella: The Neural Response When Telling Apart What is Real from

What is Fictional in the Context of Our Reality." *Journal of Cognitive Neuroscience.*

20(6):965-976. Retrieved May 20, 2021 (https://doi.org/10.1162/jocn.2008.20059).

Beckwith, Sarah. 2015. "Are There any Women in Shakespeare's Plays?" *Johns Hopkins*

University Press. 46(2). Retrieved April 25, 2021 (http://doi.org/10.1353/nlh.2015.0018).

Cantavella, Anna Juan. 2017. "Between Fiction and Reality: Maps and Cartographic Logic in the

Works of Peter Sís." *Children's Literature in Education.* 48(39-55). Retrieved April 25,

2021 (https://doi-org.ezproxy.library.wwu.edu/10.1007/s10583-016-9304-4).

Centers for Disease Control and Prevention. (2017-2018). *National Health and Nutrition*

Examination Survey. Retrieved from https://wwwn.cdc.gov/nchs/nhanes/

continuousnhanes/default.aspx?BeginYear=2017.

Dill-Shackleford, Karen E., Cynthia Vinney, and Kristin Hopper-Losenicky. 2016. "Connecting

the dots between fantasy and reality: The social psychology of our engagement with

fictional narrative and its functional value." *Social and Personality Psychology Compass.*

10(11):634-646. Retrieved May 20, 2021 (https://doi-org.ezproxy.library.wwu.edu/

10.1111/spc3.12274).

Hill, Colin. 2012. *Modern Realism in English-Canadian Fiction.* University of Toronto Press.

Jacobs, Arthur M., and Roel M. Willems. 2018. "The Fictive Brain: Neurocognitive Correlates of

Engagement in Literature." *Review of General Psychology.* 22(2):147-160. Retrieved

May 20, 2021 (https://doi-org.ezproxy.library.wwu.edu/10.1037%2Fgpr0000106).

NHANES. "National Health and Nutrition Examination Survey: Overview." *National Center for*

   Health Statistics. Retrieved May 9, 2021 (https://www.cdc.gov/nchs/data/nhanes/

   nhanes_13_14/NHANES_Overview_Brochure.pdf).

Orthia, Lindy A. 2019. "How Does Science Fiction Television Shape Fans' Relationships to

   Science? Results from a Survey of 575 'Doctor Who' Viewers." *JCOM: Journal of*

   Science Communication. 18(4). Retrieved May 20, 2021 (https://doi.org/

   10.22323/2.18040208).

Tartt, Donna. 1992. *The Secret History.* New York: Vintage Contemporaries.