Genes and Sales

Shiyang Gong
Associate Professor
Business School, Beijing Normal University
gongshiyang@bnu.edu.cn

Qian Li
Professor
International Business School, Beijing Foreign Studies University
liq@bfsu.edu.cn

Song Su
Professor
Business School, Beijing Normal University
sus@bnu.edu.cn

Juanjuan Zhang
John D. C. Little Professor of Marketing
MIT Sloan School of Management, Massachusetts Institute of Technology
jjzhang@mit.edu

March 12, 2023

Acknowledgements: The authors thank the editor, associate editor, and reviewers for their support and excellent suggestions. The authors received many helpful comments from Remi Daviet, Xing Li, Gideon Nave, Jing-Ke Weng, and Yingjie Zhang; from seminar participants at the Asia School of Business, City University of Hong Kong, Duke University, East China Normal University, Harvard University, Hong Kong Polytechnic University, Leshan Normal University, Massachusetts Institute of Technology, Peking University, Renmin University, Sun Yat-sen University, Temple University, University of Houston, University of Pennsylvania, University of Science and Technology of China, University of Western Ontario, and Xiamen University; and from attendees of the 2021 ISMS Doctoral Consortium and 2021 Marketing Modelers' Meeting. Zijing (Jimmy) Hu, Yuexin Song, and Weichao Yang provided outstanding research assistance.

Funding: This work was supported by grants from the National Natural Science Foundation of China (Nos. 71972040, 72072014, 71872016, 71702010, 71602033, and 71572018).

Genes and Sales

Abstract

This paper presents one of the first marketing applications of modern genetics. We report evidence that salespeople's genetic variants linked to educational attainment predict sales performance. Both genetics and selling effort contribute to sales performance, whereas genetics contribute more than personality traits. We further show that adaptive learning, as captured in salespeople's customer orientation and opportunity recognition skills, may explain the gene-sales relationship. We discuss the implications of these findings for sales management and the value of genetic research for the marketing field.

Keywords: genetics, molecular genetics, polygenic score, sales management, adaptive learning

1 INTRODUCTION

Sales is by far the most expensive marketing function. Ahearne (2019) reports a 1.2 trillion-dollar annual spending on sales among North American companies, which is seven times the amount of spending on traditional marketing and 33 times the amount for digital marketing. The return can be enormous. Onboarding the right salespeople multiplies company revenue (Weinfurter 2017). However, it is famously challenging to assemble, develop, and retain an effective sales force (Mayer and Greenberg 1964, Sunder et al. 2017, Williams 2017).

The challenge is real for salespeople themselves. With a constant need to communicate, persuade, and overcome rejection, the sales job is not for everyone (e.g., Ryan 2017). Some would even rank sales as one of the most stressful occupations.[1] The problem worsens when people do not know whether they will be good at sales while facing career choices.[2] Many would actually try the sales job, where the all-too-often lack of fit would then lead to costly attrition. Indeed, annual turnover in the U.S. sales industry is as high as 27%, twice as much as the rate in the overall labor force.[3]

Behind the problem is the lack of clarity on what it takes to be a good salesperson. Practitioners lament the lack of "rigor and science" to guide sales force management, making it "ironically the least disciplined process in the business" (Weinfurter 2017). Academic researchers do not have a conclusive answer either. Even one of the most fundamental questions remains open as of today: Are good salespeople born or made (e.g., Ahearne 2019, Loveland et al. 2015)? The answer is often elusive due to the lack of data on what is considered "born."

We overcome this challenge by measuring salespeople's genetic endowment. We ask whether genetics predict sales performance and, if so, to what extent and why. We study 117 salespeople of a telemarketing company over 13 months, from August 2018 through August 2019. With informed

[1] Source: https://www.payscale.com/career-news/2011/02/most-stressful-jobs.

[2] Related discussions abound on the internet (e.g., https://www.quora.com/How-do-I-know-if-Ill-be-good-at-sales).

[3] Source: https://hbr.org/2017/07/how-to-predict-turnover-on-your-sales-team.

consent, we collect the salespeople's genetic data via samples of saliva. For sales performance, we track the industry-standard metric – the amount of revenue generated by each salesperson. Meanwhile, we obtain data on a rich ensemble of factors that may affect sales performance, including sales task characteristics, salesperson demographics, personality traits, selling effort, sales skills, parental environments, and verbal skills.

We use the "polygenic score" method to analyze the genetic data. This method leverages findings from state-of-the-art genome-wide association studies (GWASs), which scan the entire genome to identify genetic variants systematically associated with an observed outcome (called "phenotype" in genetics). Given the enormous number of potential variants along the human genome and the potentially tiny effect of each variant, GWASs rely on massive samples to make replicable discoveries. The polygenic score approach, on the other hand, does not require massive samples. In a way analogous to "transfer learning" (Zhuang et al. 2020), the polygenic score approach summarizes GWAS discoveries into a single index, a polygenic score, and uses the score as a measure of genetic endowment to study genetic effects in a related domain.

We construct salespeople's polygenic scores by weighing their actual genetic data with effect sizes estimated from the largest GWAS on educational attainment by the time of our study (Lee et al. 2018). We choose this GWAS because it is reliable and relevant.[4] Educational attainment is the first and prototypical phenotype studied in social-science GWASs. Its polygenetic score reliably predicts people's years of schooling in a series of GWASs of increasing sample size, surpassing one million individuals in our chosen GWAS (Lee et al. 2018). Out of sample, the educational attainment polygenic score predicts socioeconomic outcomes such as life-course development (Belsky et al. 2016) and wealth (Barth et al. 2020). Biologically, the educational attainment polygenic score is related to cognitive functions such as learning (Belsky et al. 2016, Lee et al. 2018, Okbay et al. 2016), which arguably matters for sales professionals.

[4] This is analogous to choosing ResNet as a pre-trained network for image studies, for its reliability and relevance.

As proof of face validity, we find that salespeople's polygenic scores indeed predict their actual educational attainment, replicating earlier findings from the literature. For our main question, salespeople's polygenic score predicts their sales performance. This is true even after we control for actual educational attainment, as well as task characteristics and other demographic variables. This result is robust when we recalculate the polygenic score using weights from another landmark GWAS on educational attainment (Okbay et al. 2016). This result also survives falsification tests using polygenic scores related to height (Wood et al. 2014) and waist-hip ratio (Shungin et al. 2015) as measures of genetic endowment – these physical traits are not expected to significantly affect telemarketing performance.

In further analysis, we compare the explanatory power of salespeople's polygenic score with two common drivers of sales performance – selling effort and personality traits. Both the polygenic score and effort contribute to sales performance, and contribute independently with no interaction effect. This result suggests that good salespeople may be both born and made, as greater selling effort may offset a salesperson's natural disadvantages. Moreover, what is "natural" goes beyond personality; the polygenic score explains sales performance more than personality traits.

Our mechanism analysis, although exploratory, sheds some light on the possible reason behind the gene-sales relationship. We uncover a partial mediating variable, salespeople's adaptive learning skills, as measured by their customer orientation and opportunity recognition tendencies. These results echo earlier findings that the educational attainment polygenic score is associated with cognition and learning (e.g., Belsky et al. 2016, Lee et al. 2018, Okbay et al. 2016). These findings are also intuitive; in the telemarketing context, success often relies on the salesperson being able to understand the customer and adapt to new information in real time. We confirm this intuition via semi-structured interviews with company management. Finally, we find no evidence of parental environments, on-the-job-learning, or verbal skills explaining the gene-sales relationship.

Our findings can be substantively useful in several ways. We find that "nature" and "nurture" are both important in the sales profession. Knowing the nature of sales performance can also inform

actions to better nurture the sales force. For individuals who wish to make better-informed work and career decisions, genetics can serve as an additional information source.[5] For companies, our findings suggest that incentive systems should be designed to motivate heterogeneous salespeople. During sales force development, greater focus on adaptive learning skills can enhance productivity. Alternatively, redesigning sales tasks to lower the pressure of adaptive learning can reduce productivity gaps in the sales force. Finally, it may be worthwhile to base sales compensation on not only outcome but also effort.

Methodologically, our study represents one of the first marketing applications of modern genetics (see Daviet et al. 2022 for a review). In particular, our paper highlights the value of the polygenic score approach as a way to study human genetics in business contexts. We collect primary genetic and business outcome data while leveraging genetic effects robustly estimated from massive-scale GWASs. Many GWASs today rely on archival data that have accumulated through many years of research collaborations around the globe. These databases offer the sample size GWASs require, but rarely track business outcomes such as sales performance (Friedman et al. 2021). The polygenic score approach allows researchers to overcome sample-size constraints to study novel, managerially important, and application-driven outcomes. Besides its business value, the inclusion of more phenotypes in the human genetic knowledge base is a contribution to science in itself.

2 METHODOLOGICAL BACKGROUND

We begin by presenting background information on the human genome, behavioral genetics methodologies, and the polygenic score for educational attainment. We keep the presentation intuitive for readers new to the topic. We refer interested readers to Beauchamp et al. (2011), Benjamin et al. (2012), Friedman et al. (2021), and Uffelmann et al. (2021) for further details.

[5] We surveyed 500 individuals on MTurk, asking: "If there is a genetic test to predict if you will be a good salesperson, would you be interested?" Among those who responded, 79 (16.5%) said no, 158 (32.9%) said somewhat, and 243 (50.7%) said yes.

2.1 The Human Genome

A human cell contains 23 pairs of deoxyribonucleic acid (DNA) molecules called chromosomes. Each pair inherits one copy of a chromosome from each parent of the individual. A DNA molecule is a double-helix structure, stabilized by nucleotide base pairs of two possible types: the adenine (A)-thymine (T) pair or the cytosine (C)-guanine (G) pair; see Figure OA1 in the Online Appendix for an illustration. The human genome contains about 3 billion base pairs, the permutation of which determines the construction and operation of the human body.

Approximately 99% of locations along the human genome feature identical base pairs among all humans and thus do not reflect differences across people. At the remaining approximately 1% of genome locations, inter-individual variations in base pairs, called polymorphisms, can be observed. The most common type of polymorphisms is the single-nucleotide polymorphism (SNP, simply pronounced as "snip"), which refers to variation in the base pair at a particular genome location.[6] SNPs play a vital role in the function of genes, which are DNA fragments with the code to determine protein synthesis, through regulation of the gene expression process.

For each human SNP, there are two possible base pair types: the type more frequently observed in the population is called the major allele and the less frequent type is called the minor allele. Recall that each individual inherits two copies of chromosomes from her/his parents. Therefore, each individual can have 0, 1, or 2 copies of the minor allele at each SNP location. Human genomic information is thus commonly encoded as a numerical sequence, mapping an individual's minor allele count (i.e., 0, 1, or 2) at each SNP location of her/his genome. This allows genetic researchers to quantify the association between SNPs and various phenotypes of interest.

SNP analysis is valuable for human behavioral research for several reasons. First, it helps

[6] Other forms include insertions/deletions (indels) and a diverse set of structural variants (SVs) (Feuk et al. 2006). To date, there are over 1 billion identified SNPs according to the National Center for Biotechnology Information dbSNP Build 155 (RefSNP count 1,053,623,523; source: https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi).

researchers identify SNP locations associated with human behaviors and, in doing so, predict these behaviors. While surveys and psychological tests are established methods to provide higher-level constructs that predict behaviors, SNP analysis can add a more granular set of predictors. Second, SNPs critically affect gene functioning. Over 20,000 genes have been identified in the human genome, of which a large proportion have known functions. This provides guidance for researchers to understand the mechanisms behind a given behavior. Third, SNP analysis is arguably more accurate than behavioral tests or questionnaires; it is not constrained by subjects' comprehensive ability, mood, or involvement in the task. Finally, SNP analysis is becoming increasingly feasible to conduct. For researchers interested in archival genetic data, many providers (e.g., UK Biobank, US Health and Retirement Survey) have added molecular genetic data to their lists of variables. For researchers interested in primary genetic data, participants can collect and provide their genetic data by rubbing the inside of their cheeks for a few seconds or just spitting. The costs of genotyping technologies are also declining rapidly. This increasing availability of genetic data opens up a valuable opportunity for social scientists to study the relationship between human genetics and behaviors (for recent reviews, see Daviet et al. 2022, Harden and Koellinger 2020).

2.2 Behavioral Genetics Methodologies

Behavioral genetics is the study of genetic influence on human behaviors. During the early days of behavioral genetics, a prevalent method was to compare the outcomes of identical twins and fraternal twins to quantify the magnitude of genetic effects (Polderman et al. 2015). Twin studies cannot detect the effect of specific SNPs. *Molecular genetics* addresses this issue directly. In recent decades, the accomplishment of the epic Human Genome Project (Collins et al. 2003) and the declining costs of genotyping make it progressively feasible to examine genetic effects at the molecular level. An early method in molecular genetics is known as candidate-gene studies – hypothesis-driven designs in which SNPs in certain genes are pre-selected and hypothesized to be related to outcomes of interest (Tabor et al. 2002). Online Appendix OA1.2 explains twin studies and candidate-gene studies in further detail.

Increasingly, researchers have turned to *genome-wide association studies* (GWASs) – hypothesis-free, gene-agnostic investigation of associations between individual SNPs across the entire genome and human outcomes. In contrast to candidate-gene studies, GWASs emphasize correction for multiple testing, pre-registration, and replication in independent samples. To date, over 45,000 GWASs have been performed to analyze more than 5,000 traits (Sollis et al. 2023).

GWASs are not without limitations. Besides technical challenges (Friedman et al. 2021, Uffelmann et al. 2021), to make replicable discoveries GWASs require massive sample sizes, with recent studies involving millions of participants. Fortunately, years of research collaborations around the globe have accumulated biobanks and large-scale population studies that meet this sample-size requirement (see Uffelmann et al. 2021 for a list of GWAS databases). However, phenotypes recorded in these databases are often "minimal" compared with other genetics research methods (Friedman et al. 2021). This makes it impractical to use GWASs to study many business-domain phenotypes, such as sales performance, that are often absent from existing gene databases.

The *polygenic score approach* is a way to leverage the statistical power of GWASs to explore new phenotypes with modest sample-size requirements. A standard GWAS tests a huge number of SNPs for association with a phenotype Y by estimating a separate regression for each of these SNPs:

$$Y_i = a \cdot X_i + b_j \cdot SNP_{ij} + e_{ij}, \quad j \in \{1, \dots, J\}, \tag{1}$$

where Y_i is the phenotype of individual i, X_i is the vector of control variables for individual i, $SNP_{ij} \in \{0, 1, 2\}$ denotes individual i's minor allele count at SNP location j, a and b_j are paramters to be estimated, and e_{ij} is the error term. A phenotype tends to be associated with numerous SNPs, each SNP contributing a small effect. Therefore, a polygenic score is often calculated to capture the aggregate effect of SNPs on the phenotype. In its common form, an individual's polygenic score is a linear combination of her SNPs weighted by their corresponding effect sizes estimated from the GWAS and adjusted for cross-SNP correlations (see Section 4 for technical details):

$$Polygenic\ Score_i = \sum_{j=1}^{J} \tilde{b}_j \cdot SNP_{ij}. \quad (2)$$

A polygenic score is thus a univariate index of an individual's genetic predisposition for phenotype Y. Importantly, once reliably estimated from a GWAS sample, the effect size parameters ($\tilde{b}_j$) can be applied out of sample (on a new set of SNP_{ij} data) to construct this new sample's polygenic scores for phenotype Y.[7] This polygenic score can then be used as a standard variable to test a plethora of hypotheses on the new sample, including hypotheses about new phenotypes (Friedman et al. 2021). The sample-size requirement is the same as that needed to achieve standard statistical significance. This property allows researchers to feasibly collect primary genetic data to test their relationship with new phenotypes, an effort that helps expand the discovery of genetic associations.

Table OA1 of the Online Appendix summarizes the aforementioned behavioral genetics methods. We take the polygenic score approach. We collect primary genetic data of a sample of salespeople. We weigh their SNPs with GWAS-based effect sizes to construct a polygenic score. We then study the association between this polygenic score and sales performance. In the next section, we explain which GWAS to choose to construct the polygenic score.

2.3 Polygenic Score for Educational Attainment

The polygenic score approach is analogous to transfer learning (Zhuang et al. 2020), in that it leverages knowledge from a source domain (a GWAS) to facilitate discovery with lower sample-size requirements in the target domain (e.g., sales performance). For this approach to perform well, the source domain must be statistically reliable and theoretically relevant to the target domain (Nofal et al. 2018). In this section, we explain why our choice of GWAS, the educational attainment GWAS of Lee et al. (2018) meets both requirements.

[7] As part of global scientific collaboration on genetics research, there is a trend towards publicizing GWAS summary statistics; see Uffelmann et al. (2021) for a list of such databases.

First, GWASs on educational attainment, Lee et al. (2018) in particular, are famously well-powered, replicable, and robust, which is important to construct credible polygenic scores. For long, molecular genetics applications have focused on medical and health-related phenotypes. In 2013, researchers conducted the seminal GWAS of a social-science phenotype – educational attainment (Rietveld et al. 2013). The study analyzed millions of SNPs in a sample of over 100,000 individuals and identified replicable SNPs associated with educational attainment, measured as years of schooling. Subsequent studies have progressively extended the sample size to eventually exceed one million individuals (Okbay et al. 2016, Lee et al. 2018). The GWAS we draw on, the educational attainment GWAS by Lee et al. (2018), features a discovery sample of over 1.1 million individuals and identifies more than 1,000 genome-wide-significant SNPs.[8] These seminal GWASs have promoted the use of genetics in economics and psychology (Harden and Koellinger 2020). A growing literature shows that educational attainment polygenic scores based on these GWASs predict important socioeconomic outcomes out of sample, including life-course success (Belsky et al. 2016, 2018), labor earnings (Papageorge and Thom 2020), and wealth at retirement (Barth et al. 2020).

Second, GWASs on educational attainment are likely relevant to the sales setting. Substantial research suggests that genetic variants linked to educational attainment may influence biological processes related to cognitive functions such as learning. Biological pathway analyses by Okbay et al. (2016) and Lee et al. (2018) consistently find that SNPs most strongly associated with educational attainment implicate genes involved in brain development and neural function. Children with higher polygenic scores for educational attainment are also found to perform better on cognitive tests and exhibit a more rapid pace of cognitive development (Belsky et al. 2016). Because the sales job often requires learning, such as learning of customer needs and selling opportunities, it is reasonable to conjecture that the educational attainment polygenic score can be a composite genetic predictor of sales performance.

[8] This polygenic score can explain 11-13% of the variation in years of schooling (Lee et al. 2018).

Granted, there are potentially other polygenic scores relevant to the sales setting. However, despite its rapid development, GWAS research on social-science phenotypes is still in its infancy. The educational attainment polygenic score is arguably one of the most powerful, robust, and reliable scores in the social sciences to date (Okbay et al. 2022). Therefore, we adopt a single-polygenic score approach, focusing on the educational attainment score, rather than a multi-polygenic score approach. Including polygenic scores for other phenotypes that have not been validated and replicated by large-sample GWASs may lead to spurious or biased genetic associations.[9]

3 DATA

To implement the polygenic score approach, we obtained primary data on salespeople at a telemarketing company in Asia. In this section, we first introduce the company and its sales function. We then describe our data collection process.

3.1 The Telemarketing Company and Its Sales Function

We collaborated with a leading telemarketing company in Asia. The company's primary business was to provide outbound call services to sell personal financial products. To ensure comparability of sales performance across salespeople, we focused on one of the companies' major functions, in which salespeople had the same mission of selling installment-payment plans to credit card users. A total of 139 salespeople across two teams served in this function, 77 from "Team A" and 62 from "Team B." We included both teams in our study to increase sample size and will control for potential differences between them in subsequent analysis. Offices for both teams were located in modern business buildings, where each salesperson was assigned to a standard cubicle equipped with the same computer, telephone, and headset. Each salesperson must accomplish two weeks of pre-employment standardized training, which covered communication skills, product information,

[9] For example, among the more cited personality GWASs, Lo et al. (2017) identify 6 loci from 123,132 to 260,861 individuals, and Nagel et al. (2018) find 136 loci from 449,484 individuals. Because these studies have not been replicated in larger samples and the number of identified loci varies greatly, the polygenic scores calculated from these GWASs usually do not yield robust findings in empirical studies (e.g., Tao et al. 2022).

work protocols, and simulation exercises. Figure OA2.1 in the Online Appendix shows photos of the telemarketing teams, their business buildings, and their pre-employment training.

The salespeople typically worked five days a week in the office from 9 a.m. to 5 p.m.. Each day, a salesperson on shift would log on to the company's IT system and make calls to accomplish the sales tasks given by the system. To determine the sales task, the company's IT system would assign each salesperson into one of three "task tiers" for each month based on the salesperson's performance in the previous month. Better-performing salespeople were assigned into higher tiers, which included tasks that were more difficult but would generate higher revenue if accomplished. Importantly, within each task tier, the assignment of individual sales tasks was randomized and was determined by the company's algorithm each morning of work.

The monthly earnings of a salesperson consisted of a flat salary and a bonus. The flat salary depended on seniority, education, and prior work experience, averaging around $300 per month. The bonus averaged about $520 per month and was largely a linear function of sales performance (details to follow). Salesperson promotion was also partially based on sales performance.

3.2 Data Collection

To study the gene-sales relationship, we collected primary data on salespeople's genetic variants (e.g., SNPs). We obtained archival data on actual sales performance and factors that might affect sales performance, including sales task characteristics and salesperson demographics. Further, to assess the explanatory power and explore the potential mechanism, we collected archival data on salespeople's effort, questionnaire data on their personality traits, sales skills, and parental environments, and lab data on their verbal skills. We also interviewed company management to gain further insight into the mechanism. The archival data were directly extracted from the company's central database. We manually collected the genetic data, questionnaire data, lab data, and interview data. We present the genetic-data collection process below. We detail the questionnaire, lab, and interview data and their collection in subsequent sections.

We collected the genetic data on November 22, 2018 and March 14, 2019 for the two telemarketing teams, respectively. The data collection processes were identical for both teams. One week before data collection, each salesperson from the focal team received an invitation to our study, with details about its objectives and the procedures. All 139 salespeople in the two teams agreed to participate. Each of them received a gift worth $10 for participation following company suggestion.

Data collection proceeded as follows (see photos in Figure OA2.2 of the Online Appendix). All participants were invited into a separate office provided by the company and were informed of the purpose of the study and the use of their data. First, with informed consent, we collected the saliva sample from each participant. Under the guidance of research assistants, each participant received a DNA test kit and followed the instructions to provide a saliva sample. The research assistants checked each sample for quality and helped transfer the samples promptly to a certified DNA testing laboratory. DNA extraction and SNP genotyping were subsequently performed at the laboratory, which charged around $200 per person to process the DNA data.

The study protocol received university ethical review board approval. Participants in the study were well-informed and gave written consent before their data were collected.

4 DATA PROCESSING AND DESCRIPTIVE STATISTICS

In this section, we explain the methods we used to process the genetic data and construct the polygenic score. We then present the resulting sample for subsequent analysis and its descriptive statistics. These methods involve many technical choices. We provide references throughout this section for readers interested in their scientific foundation.

4.1 Genotyping, Quality Control, and Sample

A certified DNA testing laboratory performed DNA extraction and genotyping on saliva samples from the 139 salespeople in our data. An Illumina platform (Illumina HumanOmniZhongHua-8 chip, version 1.3) was used to assay common SNP variations in their genomes.

We implemented standard quality control of the genotype data using the PLINK software (version 1.9), including checks for individual missingness, sex discordance, heterozygosity rate, relatedness, SNP missingness, minor allele frequency, and Hardy-Weinberg equilibrium.[10] The purpose of genotype data quality control is to identify and filter out individuals and SNPs that may introduce bias to subsequent genetic analyses (see Online Appendix OA3 for further details). This process removed 22 such unqualified individuals from our original data (Anderson et al. 2010).

Table 1: Summary Statistics

Variable	Mean	S.D.	Min	Max
Cross-Sectional Data (N = 117)				
Age	22.838	3.739	18	41
Female	0.632	0.484	0	1
Tenure	0.641	0.688	0	3
Team	0.581	0.495	0	1
Years of education	13.479	1.750	9	16
Highest education				
Middle school	0.034	0.182	0	1
Secondary specialized school	0.402	0.492	0	1
High school	0.051	0.222	0	1
Advanced specialized school	0.470	0.501	0	1
College	0.043	0.203	0	1
Panel Data (N = 1,053)				
Sales performance ($1,000)	1.533	1.331	0.014	11.368
Task tier				
Low	0.356	0.479	0	1
Medium	0.363	0.481	0	1
High	0.281	0.450	0	1

Notes. An observation in the cross-sectional data is a salesperson. An observation in the panel data is a salesperson-month combination. Sales performance is monthly revenue divided by #working days of a month ($1,000). Age and tenure are in years. Female is 1 for female and 0 for male. Team is 1 for telemarketing Team A and 0 for Team B.

The resulting sample includes 117 salespeople. The top panel of Table 1 presents their summary

[10] Following well-established guidelines (Anderson et al. 2010, Marees et al. 2018), we used the following criteria: individual missingness < 0.07, X chromosome homozygosity estimate > 0.8 (male) and < 0.2 (female), heterozygosity rate ± 3 SD of the mean, identity by descent (IBD) < 0.1875, SNP missingness < 0.05, minor allele frequency (MAF) > 5%, and Hardy-Weinberg equilibrium p-value > 0.000001.

statistics. Their average age was 22.8 years and 63.2% were female. They had been with the company for an average of 0.64 years. They had received 13.5 years of education on average. The breakdown of their highest education was 3.4% middle school, 40.2% secondary specialized school, 5.1% high school, 47.0% advanced specialized school, and 4.3% college.

For each salesperson in our sample, we obtained longitudinal data on several variables, which allowed us to assemble a panel dataset for subsequent analysis. Specifically, for the outcome variable, we used the industry-standard objective measure of sales performance: daily revenue generated by a salesperson.[11] The company shared the amount of revenue generated by each salesperson in the sample for each month from August 2018 through August 2019. A salesperson's average daily revenue of a month was then calculated as her/his monthly revenue divided by the number of working days of that month. We also tracked each salesperson's task tier for each month, which was assigned by the company depending on past performance (Section 3.1).

As a result, our primary data for sequent analysis was a 13-month panel dataset including 117 salespersons and 1,053 salesperson-month observations.[12] The bottom panel of Table 1 presents the summary statistics of this panel dataset. A salesperson generated $1,533 of daily revenue on average. Over the study period, 35.6% of salesperson assignments were in low-tier tasks, 36.3% medium-tier tasks, and 28.1% high-tier tasks.

4.2 Genotype Imputation

Genotype imputation is a process of inferring missing genotypes that are not directly assayed in a study sample (Marchini and Howie 2010). By extrapolating genetic correlations from a densely genotyped reference panel to a more-sparsely genotyped study sample, imputation can estimate unobserved SNPs, thereby expanding SNP coverage and increasing the chances of finding true

[11] Revenue was calculated by multiplying a customer's deal amount with installment rate, which ranged from 3% to 18% depending on the customer's repayment plan.

[12] Due to high turnover rate, not all 117 salespeople worked for all 13 months during our study period. A salesperson was present for an average of 9 months (SD = 3.25). Thus, we analyzed an unbalanced panel dataset in this paper.

associations in a genetic study (Howie et al. 2009).

Genotype imputation has emerged as a standard practice in genetic studies for its cost effectiveness and proven accuracy. Modern genetic studies often require extremely large samples of individuals, which may escalate the cost of genotyping. A prevailing solution is to use a lower-cost and reduced-density SNP array with markers optimized for imputation, leveraging information from dense reference panels (Phocas 2022). Many studies have shown that imputation produces polygenic scores that are highly correlated with scores derived from gold-standard whole-genome-sequencing data (e.g., Gilly et al. 2019, Homburger et al. 2019, Pasaniuc et al. 2012).

We followed one such array-based SNP genotyping method, combined with imputation to infer the full set of genotypes for polygenic score calculation (Chen et al. 2020). We present details of our imputation procedures in Online Appendix OA4 and outline the key steps below. For our data, we imputed additional SNPs using the 1,000 Genomes Phase 3 reference panel (Auton et al. 2015). Imputation was performed on all autosomal SNPs (meaning SNPs in chromosomes 1 through 22) that appear in the Single Nucleotide Polymorphism Database (dbSNP, version 151).[13] Following established guidelines (Van Leeuwen et al. 2015, Verma et al. 2014), we used the SHAPEIT2 software (version v2.r900) for phasing, followed by the IMPUTE2 software (version 2.3.2) for imputation. We identified and filtered poorly imputed SNPs by post-imputation quality control. SNPs in chromosomes X, Y, and mitochondria were also excluded because standard polygenic scoring analysis focuses on autosomal SNPs.

The resulting genetic data included a total of 5,452,987 genotyped and imputed SNPs. Table OA4 in the Online Appendix presents an overview of the genetic data. For each SNP on chromosomes 1 through 22, we present its identifier, minor allele, major allele, and allele frequency.

4.3 Polygenic Scoring

[13] The complete data for dbSNP version 151 are available at https://ftp.ncbi.nlm.nih.gov/snp.

We constructed an educational attainment polygenic score for each of the 117 salespeople in our sample by weighing each salesperson's SNP data with adjusted SNP effect sizes estimated from Lee et al. (2018). The purpose of the adjustment was to capture correlations between nearby SNPs, a phenomenon known as linkage disequilibrium (LD). There are many ways to perform the LD adjustment. We followed the most commonly used clumping and thresholding method (Choi et al. 2020), which has performed well in many studies (e.g., Belsky et al. 2019, Purcell et al. 2009).

More specifically, we followed Choi et al. (2020) to calculate polygenic scores using the PRSice software (version 2.3.3).[14] We matched genotypes from our data with GWAS results reported by Lee et al. (2018) and used the 2,827,305 matched SNPs to score each salesperson's genetic predisposition for educational attainment. We used all matched SNPs to compute the polygenic score. This practice has been shown to produce greater predictive power than including only SNPs that reach genome-wide significance (e.g., Okbay et al. 2016, Purcell et al. 2009).[15] SNPs were clumped (i.e., prioritizing SNPs at the location with the smallest GWAS *p*-value) so that the retained SNPs were largely independent of each other and, therefore, their effects could be additive. We standardized the calculated score to have a mean of zero and a standard deviation (SD) of one. The score was approximately normally distributed (see Online Appendix Figure OA9.1), consistent with findings in the literature (e.g., Barth et al. 2020, Lee et al. 2018, Okbay et al. 2016).

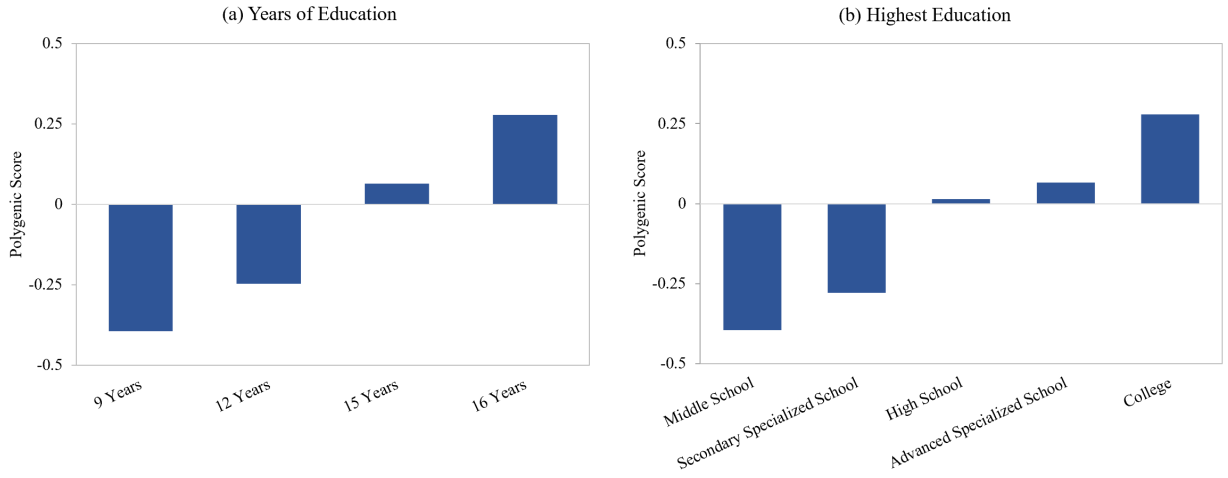
Figure 1 displays salespeople's mean polygenic scores by their actual educational achievement. Our results replicated discoveries about the genetics of educational attainment in the GWAS literature: individuals with higher polygenic scores indeed tend to have more years of education ($r = 0.202$, $p < 0.001$; Figure 1a) and achieve higher educational degrees ($r = 0.218$, $p < 0.001$; Figure 1b). These correlations were consistent with the estimates from the latest GWAS by the time of our study (Lee et al. 2018) as well as previous GWASs on educational attainment (Belsky et al.

[14] The PRSice software and its guideline are available at http://www.prsice.info/.

[15] Genome-wide significance is a specific threshold for determining the statistical significance between a SNP and a given trait. The benchmark threshold is $p < 5 \times 10^{-8}$, which is based on Bonferroni correction for all independent common SNPs across the human genome (Dudbridge and Gusnanto 2008).

2016, Rietveld et al. 2013). Given the well-documented robustness of the educational attainment polygenic score, these results suggest good face validity of our genotype data analysis thus far.

Figure 1: Association between the Polygenic Score and Actual Educational Attainment



Notes. The figure plots mean polygenic score by salesperson years of education (a) and highest education (b).

4.4 Population Stratification

Another important issue in working with genetic data is to control for population stratification – genetic differences in SNP allele frequencies across populations due to systematic differences in ancestry (Freedman et al. 2004). Population stratification may confound associations between genotype and the outcome of interest in a genetic study. For example, if a particular SNP variant is more common in a specific ancestry group, an observed association between the SNP and the outcome may reflect not only the biological function of the SNP but also common environmental factors (e.g., culture or social norms) shared by this ancestry group. When both SNP allele frequencies and environmental factors differ across populations, failure to control for these differences can lead to omitted-variables bias and thus spurious associations between these alleles and the outcome of interest.

We collected information on the ethnic origin of salespeople in our data. All of them are East Asian descents and belong in the same major ancestral group (i.e., Han ethnicity). However, even within a single ethnic group, subtle degrees of population stratification may exist (Abdellaoui et al. 2013,

2019). To control for potential population stratification, we used a standard approach – the EIGENSTRAT method developed by Price et al. (2006). This method applies principal-component analysis to the genotypic data to obtain the loadings of each individual on the top K principal components. The first principal component of a set of genotypes is the linear combination of the genotypes with the coefficients chosen to capture as much of the sample variation as possible. The second principal component is obtained in a similar manner, aiming to capture as much of the remaining variation after applying the first component. Subsequent components are obtained similarly. These principal components have been shown to capture common variation across the population and therefore serve as good controls for ancestral differences (Novembre and Stephens 2008, Price et al. 2006). How many components need to be included depends on the population structure and sample size, but including 10 is generally accepted (Beauchamp et al. 2011).

In light of the above, we included the top 10 principal components we computed using the PLINK software (version 1.9; Marees et al. 2018) to control for potential population stratification.[16] These 10 principal components together explained 9.45% of the variation in the educational attainment polygenic score. The remaining variation may come from individual-level factors such as genetic variations within the same population. For a robustness test, we controlled for different numbers of principal components (i.e., 15, 20, and 30). We obtained similar results (see Section 5.1.1).

5 ANALYSES AND RESULTS

Our analyses proceed in three steps. First, we examine the main association between genetic factors and sales performance. We show that the polygenic score for educational attainment significantly and robustly predicts sales performance. Second, we compare the effect size and predictive power of the polygenic score with selling effort and personality – factors that are commonly believed to drive sales performance. Finally, we explore possible mechanisms that may explain the gene-sales relationship.

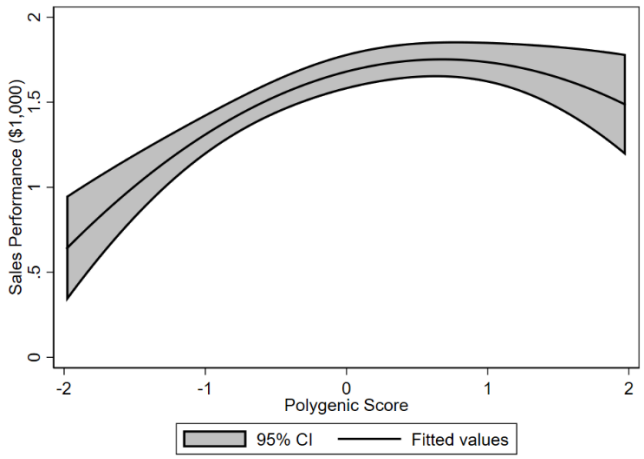
[16] See https://www.cog-genomics.org/plink/1.9/strat for details of using PLINK for principal component analysis.

5.1 Polygenic Score and Sales Performance

5.1.1 The Gene-Sales Relationship

We first present model-free evidence of the association between the educational attainment polygenic score and sales performance. Figure 2 plots the unconditional relationship between the (standardized) polygenic score and the daily sales revenue (in $1,000) generated by salespeople in our sample. The overall relationship is positive. More specifically, the relationship is positive for over 85% of the sample. It becomes flat and negative for very high values of the polygenic score, although the association is less precisely estimated because of the relatively small number of individuals in that range (see Figure OA9.1). Interestingly, Barth et al. (2020) find a similar relationship between the educational attainment polygenic score and household wealth. The curvilinear relationship also echoes many observations in the vocational psychology literature (e.g., Antonakis et al. 2017, Fine and Nevo 2008) and may suggest a "too smart for their own good" effect (Brown et al. 2021). Like Barth et al. (2020), we leave it for future research to fully understand this curvilinear relationship and focus on the average association across the spectrum of the polygenic score.

Figure 2: Association between the Polygenic Score and Sales Performance



Notes. This figure plots the unconditional relationship between the (standardized) polygenic score for educational attainment and sales performance (in $1,000), using all salesperson-month observations in our data.

Table OA9.1 of the Online Appendix presents the correlations among the polygenic score, sales performance, and a set of control variables. The polygenic score has a significant and positive correlation with sales performance ($r = 0.133$, $p < 0.001$).

To examine the gene-sales relationship more formally, we turn to regression analysis and estimate the following Ordinary-Least-Square (OLS) specification:

$$Sales\ Performance_{it} = \alpha + \beta \cdot Polygenic\ Score_i + \gamma \cdot PC_i + \theta \cdot Z_{it} + \varepsilon_{it}. \quad (3)$$

The dependent variable is salesperson i's sales performance (in \$1,000) of month t. $Polygenic\ Score_i$ is salesperson i's (standardized) educational attainment polygenic score. PC_i is a vector of salesperson i's top 10 principal component scores of the genetic data which, as discussed, control for potential population stratification. Z_{it} is a set of control variables, capturing age, gender, tenure, team, task tier, actual educational attainment, and year and month indicators. Lastly, ε_{it} is the standard normal error and α, β, γ, θ are parameters to be estimated.

Table 2 reports the estimation results of Equation (3) and its nested models. Column (1) presents the association between sales performance and the polygenic score for educational attainment, without additional control variables. For comparison, column (2) presents the association between sales performance and years of education, the standard measure of educational attainment (Lee et al. 2018). Both associations are positive and significant ($p < 0.01$). A one SD increase in the polygenic score is associated with a \$177 gain in sales performance ($\beta = 0.177$), which is equivalent to about 2.85 more years of education ($\theta = 0.062$). The comparison of R^2 between columns (1) and (2) suggests that the polygenic score has notably greater predictive power than years of education, explaining three times as much variation in sales performance.

Columns (3)-(6) build on column (1) and, in a stepwise manner, introduce control variables that may influence sales performance. Column (3) introduces age, the female indicator, tenure, the team indicator, task tier, and year and month fixed effects. Column (4) adds the top 10 principal

Table 2: Polygenic Score and Sales Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Polygenic score	0.177***		0.147***	0.134***	0.123***	0.127***
	(0.033)		(0.032)	(0.032)	(0.033)	(0.034)
Years of education		0.062***			0.047**	
		(0.023)			(0.022)	
Age			-0.009	-0.014	-0.022*	-0.016
			(0.009)	(0.010)	(0.011)	(0.013)
Female			0.052	0.109	0.074	0.083
			(0.077)	(0.077)	(0.077)	(0.077)
Tenure			0.316***	0.352***	0.359***	0.356***
			(0.072)	(0.078)	(0.077)	(0.079)
Team			0.691***	0.729***	0.737***	0.752***
			(0.088)	(0.094)	(0.094)	(0.095)
Task tier: medium			0.694***	0.681***	0.690***	0.684***
			(0.059)	(0.060)	(0.061)	(0.061)
Task tier: high			1.422***	1.411***	1.408***	1.408***
			(0.115)	(0.115)	(0.115)	(0.115)
Secondary specialized school						0.418***
						(0.155)
High school						0.313*
						(0.179)
Advanced specialized school						0.504***
						(0.157)
College						0.372
						(0.315)
Year fixed effects	No	No	Yes	Yes	Yes	Yes
Month fixed effects	No	No	Yes	Yes	Yes	Yes
Principal components	No	No	No	Yes	Yes	Yes
Observations	1,053	1,053	1,053	1,053	1,053	1,053
R^2	0.018	0.006	0.259	0.286	0.288	0.290
Mean VIF	1.00	1.00	2.59	2.12	2.12	2.79

Notes. An observation is a salesperson-month combination. The dependent variable is *Sales Performance* ($1,000). The *Polygenic Score* is calculated based on the educational attainment GWAS of Lee et al. (2018) and standardized to have a mean of zero and a SD of one. OLS estimates with robust standard errors reported in parentheses. VIF means variance inflation factor. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

components of the genetic data. Column (5) further includes the salesperson's years of education, whereas column (6) instead uses binary indicators for each salesperson's highest level of education. We do not include years of education and the highest education in the same regression to avoid

collinearity. The correlation between years of education and the highest-education categorical variable is 0.9671 ($p < 0.0000$).

We treat column (6) as our "main specification" because it includes the most control variables and achieves the highest R^2 across all columns. However, the polygenic score is significant across all specifications ($p < 0.01$). Its effect also remains economically sizable after the inclusion of control variables; a coefficient of 0.127 means that a one SD increase in the polygenic score is associated with $127 more in daily sales revenue, which is about 8.28% of the daily revenue earned by a salesperson in our sample. For a robustness test, we control for 15, 20, and 30 principal components instead. The polygenic score coefficient become 0.138, 0.137, and 0.134, respectively, all significant at the $p < 0.01$ level. We retain the specification with 10 principal components for a conservative estimate of the genetic effect.

Notably, the polygenic score is significantly associated with sales performance even after controlling for actual educational attainment. This suggests that the predictive power of the polygenic score may come from the underlying genetic effects it captures, in addition to the direct effect of actual educational attainment (Barth et al. 2020).

Among the control variables, tenure has a positive, significant coefficient. This could mean that more experience working at the company helps sales performance, or that better-performing salespeople choose to stay with the company longer. Being in higher task tiers is associated with greater sales performance. This is expected because, as discussed, more challenging tasks tend to generate greater revenue. Receiving specialized education is also associated with greater sales performance. Age and gender, however, are statistically insignificant, suggesting that the polygenic score informs sales performance beyond what is captured by these common demographic variables.

5.1.2 Robustness

We test the robustness of the gene-sales relationship with respect to alternative construction of the

polygenic score, using weights estimated from the second largest GWAS on educational attainment by the time of our study (Okbay et al. 2016). We repeat the specifications in Table 2 using this alternative polygenic score. The results appear in panel A of Table 3. The polygenic score coefficient remains statistically significant across all specifications ($p < 0.05$). Its effect size is smaller than its counterpart in Table 2. This is expected because Okbay et al. (2016) used a much smaller discovery sample (293,723 individuals) and identified much fewer genome-wide-significant variants (74 SNPs) compared with Lee et al. (2018). The resulting polygenic score is thus expected to have less predictive power (Benjamin et al. 2012, Friedman et al. 2021).

As a falsification test, we analyze two additional polygenic scores related to physical traits of the salespeople: height and waist-hip ratio. Salespeople in our study only contact randomly assigned customers by phone and are otherwise not in touch with the customers. Therefore, genetics for physical appearance are not expected to affect sales performance. We use published results from large-scale GWASs of human height (Wood et al. 2014) and waist-hip ratio (Shungin et al. 2015) to calculate two corresponding polygenic scores for the salespeople. As seen in panels B and C of Table 3, indeed, neither polygenic score has significant association with sales performance.

We also explore the candidate-gene approach and test the association between sales performance and three commonly examined candidate genes: COMT, BDNF, and DRD2. We present the full details in Online Appendix OA5. None of these candidate genes predicts sales performance ($p > 0.10$ for all three). These results, together with the falsification tests, confirm and complement our main finding that the polygenic score for educational attainment is especially predictive of sales performance, a finding that motivates our mechanism exploration in Section 5.3.

5.1.3 Comments on the Gene-Sales Relationship

Causality. A natural question is whether we can infer a causal relationship between the polygenic score and sales performance. On the hopeful side, genotypes are randomly assigned from parents

Table 3: Alternative Polygenic Scores and Sales Performance

	(1)	(2)	(3)	(4)	(5)
Panel A: Polygenic Score for Educational Attainment – Alternative Weights (Okbay et al. 2016)					
Polygenic score (educational attainment)	0.140***	0.102***	0.084**	0.076**	0.076**
	(0.034)	(0.031)	(0.036)	(0.036)	(0.036)
Standard controls	No	Yes	Yes	Yes	Yes
Principal components	No	No	Yes	Yes	Yes
Years of education	No	No	No	Yes	No
Highest education indicators	No	No	No	No	Yes
Observations	1,053	1,053	1,053	1,053	1,053
R^2	0.011	0.253	0.280	0.283	0.285
Panel B: Polygenic Score for Height (Wood et al. 2014)					
Polygenic score (height)	-0.076*	0.013	0.008	0.003	0.002
	(0.043)	(0.036)	(0.036)	(0.036)	(0.036)
Standard controls	No	Yes	Yes	Yes	Yes
Principal components	No	No	Yes	Yes	Yes
Years of education	No	No	No	Yes	No
Highest education indicators	No	No	No	No	Yes
Observations	1,053	1,053	1,053	1,053	1,053
R^2	0.003	0.247	0.277	0.281	0.282
Panel C: Polygenic Score for Waist-Hip Ratio (Shungin et al. 2015)					
Polygenic score (waist-hip ratio)	0.021	0.005	0.027	0.039	0.035
	(0.041)	(0.035)	(0.037)	(0.037)	(0.038)
Standard controls	No	Yes	Yes	Yes	Yes
Principal components	No	No	Yes	Yes	Yes
Years of education	No	No	No	Yes	No
Highest education indicators	No	No	No	No	Yes
Observations	1,053	1,053	1,053	1,053	1,053
R^2	0.000	0.247	0.277	0.282	0.283

Notes. An observation is a salesperson-month combination. The dependent variable is *Sales Performance* ($1,000). Panels A-C present the effect of the polygenic score based on three alternative GWASs, respectively. Standard controls include age, female, tenure, team, task tier, and year and month fixed effects. OLS estimates with robust standard errors reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

to an offspring at conception and remain mostly fixed throughout the lifespan. Therefore, an individual's polygenic score constitutes the most stable type of personal variables. The estimated effect of the polygenic score should be largely unsusceptible to the problem of reverse causality (Madole and Harden 2021). However, it is difficult to ascertain causality, a major challenge being the potential confounding of gene-environment correlations (Koellinger and Harden 2018).

Parents not only pass on genetic materials to their children but also determine their parental environments. Therefore, the polygenic score may reflect not only genetic factors that causally affect sales performance, but also parental environments that affect sales performance in their own ways (e.g., parents' own job experience in sales). If the polygenic score and environmental variables are positively correlated, omitting environmental variables will lead to an upward-bias in the estimated coefficient of the polygenic score. Reassuringly, as we will explain in Section 5.3.2, the gene-environment correlation is unlikely to be a quantitatively important confounder in our study, because the self-reported effect of parental environments by the salespeople in our sample are systematically low without sizable variation.

Explanatory power. The polygnenic score in our study explains 1.8% of the variation in sales performance. This may be a conservative quantification of the genetic effect for at least two reasons. First, since polygenic scores are constructed based on GWAS results, estimation errors in the GWAS parameters ($\tilde{b}_j$) will lead to measurement errors in the polygenic scores. Such measurement errors may in turn attenuate the association between the polygenic score and sales performance. Indeed, R^2 comparison between Tables 2 and 3 shows that our polygenic score has 64% more explanatory power than the score based on the smaller educational attainment GWAS of Okbay et al. (2016). The accuracy and explanatory power of the polygenic score will increase as GWASs become larger leading to replicable SNP effects (Harden and Koellinger 2020), which is the direction the GWAS literacture is progressing.

A second reason has to do with the functional form used in the original GWAS and in constructing the polygenic score. Equations (1) and (2) assume an additive, linear relationship between SNPs and the outcome. For more than a decade, this additive linear model has been widely used in thousands of molecular genetic studies. However, the underlying molecular networks may be nonlinear (Marjoram et al. 2014) and SNP effects may interact (Zuk et al. 2012). For these reasons, polygenic scores based on the additive linear model tend to underestimate genetic effects (Manolio et al. 2009). We followed the standard, additive linear model. Our results should therefore be seen

as a conservative test of genetic effects on sales performance.

5.2 Contribution to Sales Performance: Polygenic Score, Effort, and Personality

The results so far highlight a significant and robust gene-sales relationship. The next question regards the extent to which genetics explains sales performance compared with other established explanations. We examine two common explanations of sales performance: effort and personality.

Past research shows that one of the most prominent drivers of sales performance is salesperson motivation (Churchill et al. 1985, Mayer and Greenberg 1964, Verbeke et al. 2011). In the sales domain, motivation is viewed as the amount of effort a salesperson desires to expend on the activities associated with her/his job (Walker et al. 1977). Conceptually, effort contrasts saliently with genetics because a salesperson can control her/his effort but not genetic endowment. For a proxy measure of selling effort, we collect data on the amount of time (in hours) a salesperson spent on the phone with clients each day. In our sample, salespeople's average time spent on the phone was 3.611 hours per day (SD = 0.438).

Past research also shows that personality traits can predict job performance in many job categories; sales jobs are no exception (e.g., Barrick and Mount 1991, Salgado 1997). According to "trait activation" theory (Tett and Burnett 2003), which studies situational specificity of personality-job relationships, personality traits matter particularly when salespeople try to persuade others. We use the well-established NEO Five-Factor Inventory (Costa and McCrae 1992) to measure each salesperson's Big Five personality traits: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness (see Online Appendix OA6 for details). We calculate the score of each personality dimension as the mean value across its corresponding items (negative-phrased items are reverse-coded). Cronbach's α, a measure of scale reliability, is above 0.7 for each dimension. Cronbach's α of the overall scale is 0.899.

Table OA9.2 of the Online Appendix presents the summary statistics and correlation matrix among

sales performance, the polygenic score, selling effort, and the five personality traits. Notably, the correlation between the polygenic score and selling effort is insignificant ($r = 0.013$), suggesting that the more genetically endowed salespeople may not work harder or less hard systematically.

Table 4: Contribution to Sales Performance: Polygenic Score, Effort, and Personality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Polygenic score	0.123***	0.130***	0.127***	0.146***	0.116***	0.122***	0.137***	0.133***
	(0.034)	(0.034)	(0.034)	(0.035)	(0.035)	(0.033)	(0.036)	(0.037)
Selling effort	0.654***							0.628***
	(0.103)							(0.105)
Neuroticism		-0.052					0.017	0.011
		(0.041)					(0.046)	(0.045)
Extraversion			0.049				-0.036	-0.026
			(0.043)				(0.055)	(0.054)
Openness to experience				0.121***			0.118**	0.114***
				(0.041)			(0.047)	(0.043)
Agreeableness					0.067		0.020	0.036
					(0.047)		(0.057)	(0.056)
Conscientiousness						0.135***	0.130***	0.084*
						(0.039)	(0.048)	(0.048)
Standard controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Principal components	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Highest ed. Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,053	1,053	1,053	1,053	1,053	1,053	1053	1,053
R^2	0.328	0.291	0.291	0.296	0.292	0.297	0.303	0.337
Mean VIF	2.77	2.77	2.76	2.76	2.79	2.77	2.77	2.75

Notes. An observation is a salesperson-month combination. The dependent variable is *Sales Performance* ($1,000). The *Polygenic Score* and each personality trait are standardized to have a mean of zero and a SD of one. Standard controls include age, female, tenure, team, task tier, and year and month fixed effects. OLS estimates with robust standard errors reported in parentheses. VIF means variance inflation factor. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

We extend our main specification by including selling effort and (standardized) personality traits as covariates, individually and then simultaneously. Table 4 shows the results, which are qualitatively and quantitatively comparable across all columns. We focus on column (8), the model with all covariates, for interpretation. Even after controlling for selling effort and personality traits, the polygenic score retains its significance and effect magnitude ($\beta = 0.133$, $p < 0.01$), similar to its baseline estimate in column (6) of Table 2. As expected, selling effort is a significant predictor

of sales performance ($\theta = 0.628$, $p < 0.01$). A one SD increase in the polygenic score is equivalent to about 11 minutes more of effort per day, or 5% of a salesperson's average effort.

Among the five personality traits, openness to experience and conscientiousness emerge as two significant predictors of sales performance ($p = 0.009$ and 0.08, respectively), which is consistent with previous findings on personality and job performance (e.g., Barrick and Mount 1991, Hurtz and Donovan 2000). A one SD increase in openness to experience and conscientiousness is associated with a $114 and $84 increase in daily sales revenue, respectively. Both personality traits have smaller effects than the polygenic score.

We also explore the interaction effects between the polygenic score and effort as well as personality traits. As shown in Online Appendix Table OA9.3, the only significant interaction is the negative moderation effect of neuroticism on the polygenic score ($\theta = -0.097$, $p = 0.016$). Of particular interest is the insignificant interaction between the polygenic score and effort, suggesting that their effects on sales performance may occur through independent channels.

In terms of explanatory power, the polygenic score, effort, and personality each explain 1.8%, 4.7%, and 0.48% of the variation in sales performance. As discussed, our study may be a conservative test of the genetic effect. Therefore, we cannot conclude whether genetics predict sales performance more or less than effort. What we can conclude is that good salespeople are likely both born and made. Genetics matter, but greater effort can offset a salesperson's natural disadvantages. Further, personality, a commonly used predictor of sales performance and arguably a less costly metric to obtain, cannot replace the polygenic score in predicting sales performance. There may be information in the polygenic score that personality, as well as selling effort, cannot fully capture. We turn to this issue in the following section.

5.3 Mechanism Exploration

5.3.1 Adaptive Learning

A natural question at this point is what mechanism may explain the gene-sales relationship. Because the polygenic score is linked to educational attainment, one possibility is that the score may reflect genetic factors that facilitate learning in the sales profession. As biological evidence of this possibility, Okbay et al. (2016) find that SNPs associated with educational attainment are disproportionately found in genomic regions that regulate gene expression in fetal brain and neural development. Lee et al. (2018) further show that these SNPs implicate genes involved in brain development and neuron-to-neuron communication. There is also field evidence of a positive link between the educational attainment polygenic score and learning-related achievements, such as cognitive development in childhood (Belsky et al. 2016) and the learning of financial management knowledge (Barth et al. 2020). Motivated by these findings, we examine the possibility that the gene-sales relationship may reflect salesperson learning.

The importance of salesperson learning has been recognized in the sales literature. There are broadly speaking two levels of salesperson learning: learning of the basic skills to sell, such as persuasion (Ford 1983, Ford et al. 1987), and "adaptive learning" that allows the salesperson to adjust sales strategies to fit the sales situation and to leverage information gained in a specific customer interaction (Leong et al. 1989, Spiro and Weitz 1990, Weitz et al. 1986), similar to the "empathy" concept highlighted in the classic article of Mayer and Greenberg (1964).

We argue that adaptive learning is more likely to mediate the gene-sales relationship in the context we study. This is because the pre-employment standardized training likely equalizes salespeople's learning of the basic sales skill. Moreover, performing well in telemarketing relies on converting numerous strangers, each over the short duration of a phone call. It is important that the salesperson quickly adapt to each sales situation and learn each customer's needs in real time. We investigate this candidate mechanism combining quantitative and qualitative methods: we both examine a mediation model and conduct semi-structured interviews with company management to elicit first-order mechanisms in their experience.

Mediation analysis. We operationalize salesperson adaptive learning drawing on established

constructs in the sales literature: customer orientation (as opposed to selling orientation) and opportunity recognition. Customer orientation involves "interacting with customers" and trying to help customers make purchase decisions that meet their needs, whereas selling orientation involves "selling to customers" and using intensive persuasion to meet short-term sales objectives (Bagozzi et al. 2012). Opportunity recognition relies on "[making] sense of signals of change (e.g., new information about new conditions) to form beliefs regarding whether or not enacting a course of action to address this change could lead to net benefits" (Grégoire et al. 2010, p. 415). In the sales domain, opportunity recognition captures a salesperson's ability to use contextual knowledge to discover customer needs (Bagozzi et al. 2012) and has been found to drive solution selling (Bonney and Williams 2009). We posit that customer orientation and opportunity recognition may be related to the educational attainment polygenic score and may explain the gene-sales relationship. Selling orientation, to the extent that it pertains to the learning of basic sales skills, serves as a falsification test of the adaptive learning mechanism; in theory it should not mediate the gene-sales relationship.

We measure customer versus selling orientation and opportunity recognition using five-point Likert scales adapted from previous studies (Bagozzi et al. 2012, Periatt et al. 2004, Thomas et al. 2001). Online Appendix OA6 presents the details and Table OA6 provides summary statistics of the scales. We calculate each variable as the mean value across its corresponding items. The Cronbach's α values are 0.732 for the customer orientation scale, 0.770 for selling orientation, and 0.870 for opportunity recognition, all suggesting high internal consistency.

We first examine whether these sales skills are indeed related to the educational attainment polygenic score. We regress the standardized variables of customer orientation, selling orientation, and opportunity recognition on the polygenic score and our full set of control variables. Panel A of Table 5 presents the results. The three columns show, respectively, that the polygenic score has a significant positive association with customer orientation (β = 0.156, $p < 0.01$), a marginally significant negative association with selling orientation (β = -0.072, $p = 0.06$), and a significant positive association with opportunity recognition (β = 0.172, $p < 0.01$).

Table 5: Polygenic Score, Sales Skills, and Sales Performance

Panel A: Polygenic Score and Sales Skills			
Dependent variable	Customer Orientation	Selling Orientation	Opportunity Recognition
Polygenic score	0.156***	-0.072*	0.172***
	(0.032)	(0.038)	(0.028)
Standard controls	Yes	Yes	Yes
Principal components	Yes	Yes	Yes
Highest ed. Indicators	Yes	Yes	Yes
Observations	1,053	1,053	1,053
R^2	0.180	0.213	0.232
Panel B: Polygenic Score, Sales Skills, and Sales Performance			
Dependent variable	Sales Performance	Sales Performance	Sales Performance
Polygenic score	0.097***	0.132***	0.091***
	(0.033)	(0.034)	(0.035)
Customer orientation	0.191***		
	(0.031)		
Selling orientation		0.073*	
		(0.039)	
Opportunity recognition			0.207***
			(0.034)
Standard controls	Yes	Yes	Yes
Principal components	Yes	Yes	Yes
Highest ed. Indicators	Yes	Yes	Yes
Observations	1,053	1,053	1,053
R^2	0.307	0.292	0.309

Notes. An observation is a salesperson-month combination. The *Polygenic Score* and the measures of *Customer Orientation*, *Selling Orientation*, and *Opportunity Recognition* are standardized to have a mean of zero and a SD of one. Standard controls include age, female, tenure, team, task tier, and year and month fixed effects. OLS estimates with robust standard errors reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

We next examine the mediating effect of these sales skills. We include each sales-skill variable in regressions of sales performance on the polygenic score and our complete set of control variables. Panel B of Table 5 reports the results. Customer orientation has a positive significant coefficient (θ = 0.191, $p < 0.01$) and selling orientation has a positive, marginally significant coefficient (θ = 0.073, $p = 0.06$), suggesting that either orientation can boost sales performance to some extent. However, only the inclusion of customer orientation reduces the coefficient of the polygenic score, from 0.127 in the main specification to 0.097. Meanwhile, opportunity recognition has a positive,

significant coefficient (θ = 0.207, p < 0.01) and its inclusion reduces the polygenic score coefficient from 0.127 to 0.091. These results suggest that customer orientation and opportunity recognition, but not selling orientation, partially mediate the gene-sales association.

Figure OA9.2 of the Online Appendix summarizes and quantifies the effect of each potential mechanism. There are two significant paths underlying the gene-sales relationship. First, a one SD increase in the polygenic score is associated with a 0.156 unit increase in customer orientation, a unit increase of which is associated with an increase in sales performance by $191. The polygenic score's indirect effect on sales performance via customer orientation is thus 0.156 × $191 = $30, with a bootstrapped 95% confidence interval of [$16, $45]. Second and similarly calculated, the polygenic score's indirect effect on sales performance via opportunity recognition is 0.172 × $207 = $36, with a bootstrapped 95% confidence interval of [$21, $53]. The polygenic score's indirect effect via selling orientation is not significant. These findings support the hypothesis that genetic factors related to salesperson adaptive learning may explain the gene-sales relationship.

Semi-structured interviews. We interviewed managers from the telemarketing company in February 2022 to gain direct, practical insights into the gene-sales relationship. We explored managers' perceptions in three topical areas: (i) the importance of genetic endowments in the sales job, (ii) paths through which "genius" salespeople achieve high sales performance, and (iii) ways to learn sales skills. Following earlier work (Wies et al. 2019, 2022), we conducted semi-structured interviews with open-ended, non-leading questions, which allowed us to capture interviewee perspectives accurately, unconfined to a fixed range of response options (Given 2008). This approach is particularly helpful when a multitude of mechanisms potentially exist. We want the most important mechanism(s) to emerge organically from the interviews.

We interviewed two managers from the company, each from one of the two telemarketing teams. As eligibility criteria for participating in the interviews, we ensured that both managers were sufficiently senior in the company, had decision power or direct influence over the sales teams, and had extensive knowledge of the industry background. The first interviewee ("Manager A")

was female with 15 years of experience in sales-related fields, managing over 400 salespeople including Team A. The second interviewee ("Manager B") was male with 6 years at the company and was in charge of Team B. We interviewed the managers sequentially and stopped once theoretical saturation occurred or no new information emerged from the interviews. All interviews were conducted via video conferencing and lasted about 40 minutes. In what follows, we summarize the interviews with respect to the three aforementioned topics.

The importance of genetic endowments in the sales job. Both interviewees mentioned genetic endowments as a critical component of being a successful salesperson. Manager A noted: "Training and experience can only make good salespeople. Truly genius salespeople are mostly born." She added: "They are naturally receptive to new things and have a great ability to learn." Manager B said: "To become a successful salesperson, genetic factors have the largest effect [40% when asked to quantify], followed by work experience (30%), schooling (20%), and other factors (10%)."

Paths through which genius salespeople achieve high sales performance. The interviewees pointed out that ordinary salespeople are adept at handling standardized sales processes. Manager A noted: "Most salespeople are accustomed to using the persuasive words they have learned in training, or passing on to customers the features and advantages of the product that have been hammered out in advance." Manager B summarized the sales tasks on which ordinary salespeople perform well as "usual and less tough." In contrast, genius salespeople were felt to excel at handling various sales situations in a more flexible manner. Manager A said: "They are more responsive to customers' challenges. For example, when customers express doubts during communication, they are keen to figure out what customers are really concerned about, demonstrate the benefits of the product from a customer-centric perspective, and expect to reach consensus with customers." Manager B commented: "Genius salespeople learn to take their customers to heart, mind, and even soul, such that they can satisfy customers better than ordinary salespeople."

Ways to learn sales skills. According to the interviewees, company-organized training was a primary way for salespeople to learn standardized sales skills. Manager B said: "One of the most

effective ways to help new hires get up to speed is to have them listen to as many demos or pitches as possible. Our regular pitch practices provide salespeople a tool-kit of persuasive utterances that are relevant, memorable, and repeatable in their actual sales situations." Meanwhile, salespeople were encouraged to be themselves and handle various sales contexts in their own ways. Manager B noted: "We teach salespeople plenty of standardized persuasive language, yet we don't want them to interact with customers as if they were robots." Manager A stated: "Our salespeople each have their own ways of thinking, feeling, and behaving – their innate talents – that they can apply productively and uniquely. Trying to change who people are is unrealistic. We should ensure our salespeople have enough discretion in the sales process so that they can leverage their natural strengths to be successful."

Overall, the tenor of the interviews corroborates our findings: genetic endowment matters in the sales job, the more endowed salespeople tend to be more adaptive in customer interactions, and these adaptive skills do not seem to come through existing training provided at the company.

5.3.2 Alternative Mechanisms

So far, our results show that adaptive learning partly explains the gene-sales relationship. However, in panel B of Table 5, the polygenic score remains statistically significant after controlling for sales skills. Recall also that one SD increase in the polygenic score is associated with \$30 + \$36 = \$66 increase in sales performance via customer orientation and opportunity recognition combined, compared with the total effect of \$127 estimated from the main specification. These results suggest that while adaptive learning explains a good proportion (i.e., \$66/\$127 $\approx$ 52%) of the gene-sales relationship, there may be other possible mechanisms worth exploring.

Parental environments. As discussed, parental environments may confound genetic effects. Adoption studies have consistently found that children adopted into well-educated families are more likely to attend college (Björklund et al. 2006, Sacerdote 2007). There is also ample evidence in the biology literature. For example, Lee et al. (2018) find that within-family associations

between SNPs and educational attainment tend to be smaller than within- *and* across-family associations. This is possibly because within- and across-family analyses that fail to control for unobserved parental environments overestimate SNP effects. Consistent with this hypothesis, Kong et al. (2018) find that parental SNPs that are not transmitted to children are nevertheless correlated with children's outcomes, suggesting an indirect link between parents' genes and children's outcomes via the rearing environment shaped by the parents.

In the sales context, it is possible that higher polygenic scores are associated with better sales performance through advantageous parental environments. To examine this possibility, we surveyed the salespeople in our sample on their parental environments (see Online Appendix OA7 for details). The survey was conducted separately in March 2022. Due to high turnover in this industry, only 48 of the 117 salespeople were still with the company. All of them completed the survey. We were not able to match this survey with the rest of the data through (anonymous) respondent identity. This prevented us from analyzing the relationship between parental enviroments and the rest of the data. However, the survey itself was informative in several ways.

First, we collected information on the salespeople's parents' occupations, a classic objective measure of parental environments with well-documented, significant influence on individual occupational choice and development (e.g., Holland 1997, Oren et al. 2013). Among the 48 salespeople, only one reported that her father is/was a salesperson. Second, we measured parental education and investment with two questions building on the vocational psychology literature (Bryant et al. 2006): (a) "Do you think your parental or family environments help you be a good salesperson?" (1 = yes, 0 = no) and (b) "Do your parents or family invest resources (e.g., money, effort, knowledge, etc.) to help you be a good salesperson?" (1 = yes, 0 = no). A limited proportion of respondents answered yes to these questions (16.67% and 12.50%, respectively). Finally, respondents were asked to indicate how much they think parental environments in general contributed to them being good salespeople. The average value reported in our sample is 14.71%

(SD = 9.23%) out of 100%.[17] Overall, these results suggest that parental environments are unlikely to explain most of the gene-sales relationship in our study.

On-the-job learning. Another possibility related to the learning mechanism is that salespeople with higher polygenic scores adapt faster to the sales job. We regress sales performance on the polygenic score, (calendar) time, the interaction term between polygenic score and time, as well as control variables. Time has a positive coefficient ($p < 0.000$), suggesting that salespeople may learn to perform better on the job. However, the interaction between time and the polygenic score is insignificant ($p = 0.134$); salespeople with higher polygenic scores do not improve faster. We replace sales performance with task tier, an indicator of salesperson rank in the company, and find similar results. Salespeople rise in task tier over time ($p = 0.038$) but those with higher polygenic scores do not rise faster ($p = 0.841$). In other words, adaptive learning may not mean adapting to the job but rather how salespeople adapt to each customer interaction, as discussed.

Verbal skills. It is natural to suspect that salespeople with better verbal skills may have an inherent advantage. Biologically, verbal skills have been shown to correlate with genetic variants linked to educational attainment (Belsky et al. 2016). Substantively, we study the telemarketing context, in which sales are made through verbal communications over the phone. We draw on the psycholinguistics literature and assess salesperson verbal skills with two standard tasks: a rapid automatized naming task and a word list reading task (Lervåg and Hulme 2009, Torgeson et al. 1999, Zhang et al. 2012). Online Appendix OA8 presents the details. 56 salespeople from one of the two telemarketing teams participated, although the sample size is sufficient according to power analysis.[18] We find no evidence that verbal skills explain the gene-sales relationship.

---

[17] We confirmed these findings with management, who explained that most salespeople grew up in poor rural areas and were unlikely to have advantageous parental environments to support their sales career.

[18] Members from the other team did not participate due to limited availability. We used G*Power software (Faul et al. 2007) to determine the sample size required for a small-sized effect f^2 of 0.02 in a linear multiple regression design; 485 observations were required for the analysis to be powered at 95%. In our regression analysis, an observation is a salesperson-month combination. The total number of observations (551) exceeds the sample-size requirement (485).

6 DISCUSSION

Our paper contributes to the marketing field by being one of the first to apply modern genetic research methods to marketing problems (see Daviet et al. 2022 for a review). Historically, the marketing field has benefited from the introduction of tools from other disciplines. Genetics can be a powerful addition to the marketing toolkit. It helps address the classic nature-versus-nurture debate that has a range of implications. We discuss several such implications in the sales context.

6.1 Implications

Genetics research has long helped people unlock useful information about themselves, such as their susceptibility to a disease. Awareness begets precaution and facilitates planning. The same can be argued for genetic research in marketing. It provides a way for salespeople, and workers in general, to better understand their natural endowments and limitations. This information can help salespeople design their work more effectively. For example, per our findings, this information can help salespeople determine the amount of effort needed to meet a sales goal. After all, knowing their own abilities (or "types") helps individuals work better in contract theory (Laffont and Martimort 2002). Besides work planning, better knowledge of their abilities may also help individuals decide whether to pursue a sales career, where stress and failure-induced turnover is prevalent (Hu et al. 2022).

Companies can benefit from genetic research findings in several ways. If genetic information helps salespeople work more effectively, it has the potential to improve customer experience and loyalty (Huang and Sudhir 2020), workforce morale, and companies' sales force management efficiency. Moreover, our findings provide evidence of genetic heterogeneity among salespeople. Knowing whether the sales force is heterogeneous – without having to observe each salesperson's ability – can already help companies make important decisions. Examples abound in the sales literature. To name a few, Chan et al. (2014) find that firms should use group-based commission if the sales team is heterogeneous and individual-based commission if the team is homogeneous. Chen and Lim

(2017) show that average team output is the best metric to motivate effort in sales contests if salespeople are heterogeneous. Chen and Chung (2021) find that companies should reward sales teams based on maximum performance, again if salespeople are heterogeneous. Our findings provide support for management strategies tailored to motivate heterogeneous sales forces.

Our finding of genetic heterogeneity among salespeople also adds nuance to incentive design. Performance-based compensation is common in the sales industry. However, it takes more effort for the less endowed to achieve the same performance. Meanwhile, those who lag in performance do benefit from more opportunities to invest effort according to our analysis. To the extent that effort is contractible, it may therefore be productive to reward salespeople on not only performance but also effort. Consistent with this view, Rao et al. (2021) find that adding activity-based pay to standard performance-based compensation increases sales force productivity.

Finally, genetic research sheds light on why some individuals naturally perform well and therefore what can be done to enhance performance. Based on our mechanism exploration, sales force training that focuses on developing adaptive learning skills may enhance performance, if these skills can be acquired. If these skills cannot be acquired easily, companies may instead devise strategies to reduce the pressure of adaptive learning. For example, instead of randomly assigning new customers to salespeople, which is the practice at the telemarketing company we collaborate with, firms can assign similar customers to the same salesperson to reduce the pressure of learning at each customer interaction. Indeed, Hu et al. (2022) find that matching customers with salespeople improves sales outcome. In other words, knowing the nature behind sales performance can help companies nurture their sales professionals.

6.2 Caveats

The value of genetic research for marketing comes with important caveats. Genetic data contain vast amounts of information about individuals, perfectly identify individuals except for identical twins, and generate informational externalities among genetically correlated individuals such as

relatives. Therefore, genetic data should be used with extreme caution. Data collection should require informed consent, transparency of the use of data, and commitment to the scope of use. Organizations who access and analyze genetic data should safeguard the data and consider measures such as privacy-preserving computing, whereby the analyst knows the model and parameters, and the outcome (e.g., polygenetic score) is computed solely on the consumer side. Moreover, genetic research results should not be used to exploit vulnerable individuals. Daviet et al. (2022) provide a detailed discussion of the ethical and legal challenges of using genetic data for marketing. We supplement their discussion by highlighting several caveats specific to the sales context.

First, companies should refrain from using genetic information for sales force segregation or discrimination. The 2008 Genetic Information Nondiscrimination Act bars employers from using individuals' genetic information when making employment decisions. We emphasize prudent use of aggregate insight from genetic research without targeting specific individuals. Examples of such aggregate insight include, as discussed, aggregate facts that salespeople are genetically heterogenous and that adaptive learning skills may improve sales performance.

Second, more research is needed to fully understand the behavioral implications of disclosing genetic information to sales professionals themselves. For instance, it is valuable to test how knowing their genetic predisposition affects salespeople's work motivation and whether the effect depends on the extent of genetic endowment. Further, while better self-knowledge may aid decision-making, it can also generate self-fulfilling believes and social stratification. These trade-offs need to be carefully researched.

7 CONCLUDING REMARKS

Using primary genetic data, we find that genetic variants associated with educational attainment predict salespeople's performance. Genetics are more predictive of sales performance than personality traits. Meanwhile, genetics and effort both contribute to sales performance, suggesting

that good salespeople may be born *and* made. Moreover, we find evidence that adaptive learning, as reflected in salespeople's customer orientation and opportunity recognition skills, partly explains the gene-sales relationship.

To our knowledge, this paper is one of the earliest applications of modern genetics in marketing (see Daviet et al. 2022 for a review). We believe this is a research area with great potential.

First, genetic data offer a granular – even molecular – view into what is traditionally relegated as "unobserved heterogeneity" across individuals. Continued discoveries of genetic effects also guide the search for possible mechanisms. With today's technology, genetic data can be accurately obtained at increasingly affordable costs. Moreover, our paper showcases the use of the polygenic score approach, which has much lower sample-size requirements than GWASs but is able to leverage reliable findings from the ever-growing GWAS literature. This property allows researchers to realistically extend genetic analysis to marketing outcomes, which are often application-driven.

Second, answering the nature-versus-nurture question can inform many marketing decisions. It will be rewarding to extend the analysis to more-complex sales management strategies, such as sales contest (Kalra and Shi 2001, Lim et al. 2009), quota (Jain 2012), and bonus (Chung et al. 2014). It will also be rewarding to go beyond sales. After all, marketing is a field that emphasizes the role of people and genetic research provides a unique way to understand people.

Last, private genetic testing is booming.[19] Companies are gathering enormous amounts of consumer genetic data, some for commercial use, while regulations are generally lacking (Daviet et al. 2022). More academic research on the various personal, social, and market consequences of genetic data can inform policies governing the use of this fundamental ingredient of human identity.

REFERENCES

Abdellaoui A, Hottenga JJ, Xiao X, Scheet P, Ehli EA, Davies GE, Hudziak JJ, et al. (2013)

[19] Source: www.forbes.com/sites/cognitiveworld/2019/04/01/the-rise-of-genetic-testing-companies-and-dna-data-race.

Association between autozygosity and major depression: Stratification due to religious assortment. *Behav. Genet.* 43(6).

Abdellaoui A, Hugh-Jones D, Yengo L, Kemper KE, Nivard MG, Veul L, Holtz Y, Zietsch BP, Frayling TM, Wray NR (2019) Genetic correlates of social stratification in Great Britain. *Nat. Hum. Behav.* 3(12):1332–1342.

Ahearne M (2019) Rethinking marketing: Michael Ahearne. *Am. Mark. Assoc.* Retrieved (July 29, 2020), https://www.ama.org/2019/02/03/rethinking-marketing-michael-ahearne/.

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT (2010) Data quality control in genetic case-control association studies. *Nat. Protoc.* 5(9):1564–1573.

Antonakis J, House RJ, Simonton DK (2017) Can super smart leaders suffer from too much of a good thing? The curvilinear effect of intelligence on perceived leadership behavior. *J. Appl. Psychol.* 102(7):1003–1021.

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.

Bagozzi RP, Verbeke WJ, Van Den Berg WE, Rietdijk WJ, Dietvorst RC, Worm L (2012) Genetic and neurological foundations of customer orientation: field and experimental evidence. *J. Acad. Mark. Sci.* 40(5):639–658.

Barrick MR, Mount MK (1991) The big five personality dimensions and job performance: a meta-analysis. *Pers. Psychol.* 44(1):1–26.

Barth D, Papageorge NW, Thom K (2020) Genetic endowments and wealth inequality. *J. Polit. Econ.* 128(4):1474–1522.

Beauchamp JP, Cesarini D, Johannesson M, van der Loos MJ, Koellinger PD, Groenen PJ, Fowler JH, Rosenquist JN, Thurik AR, Christakis NA (2011) Molecular genetics and economics. *J. Econ. Perspect.* 25(4):57–82.

Belsky DW, Caspi A, Arseneault L, Corcoran DL, Domingue BW, Harris KM, Houts RM, et al. (2019) Genetics and the geography of health, behaviour and attainment. *Nat. Hum. Behav.* 3(6):576–586.

Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A, Conley D, Fletcher JM, Freese J, Herd P (2018) Genetic analysis of social-class mobility in five longitudinal studies. *Proc. Natl. Acad. Sci.* 115(31):E7275–E7284.

Belsky DW, Moffitt TE, Corcoran DL, Domingue B, Harrington H, Hogan S, Houts R, Ramrakha S, Sugden K, Williams BS (2016) The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychol. Sci.* 27(7):957–972.

Benjamin DJ, Cesarini D, Chabris CF, Glaeser EL, Laibson DI, Age, Gene/Environment Susceptibility-Reykjavik Study:, Guðnason V, et al. (2012) The promises and pitfalls of genoeconomics. *Annu. Rev. Econ.* 4(1):627–662.

Björklund A, Lindahl M, Plug E (2006) The origins of intergenerational associations: Lessons from Swedish adoption data. *Q. J. Econ.* 121(3):999–1028.

Bonney FL, Williams BC (2009) From products to solutions: the role of salesperson opportunity recognition Geiger S, Guenzi P, eds. *Eur. J. Mark.* 43(7/8):1032–1052.

Brown MI, Wai J, Chabris CF (2021) Can You Ever Be Too Smart for Your Own Good? Comparing Linear and Nonlinear Effects of Cognitive Ability on Life Outcomes. *Perspect. Psychol. Sci.* 16(6):1337–1359.

Bryant BK, Zvonkovic AM, Reynolds P (2006) Parenting in relation to child and adolescent vocational development. *J. Vocat. Behav.* 69(1):149–175.

Chan TY, Li J, Pierce L (2014) Compensation and peer effects in competing sales teams. *Manag. Sci.* 60(8):1965–1984.

Chen H, Chung K (2021) Increasing team performance by sharing success. *J. Mark. Res.* 58(4):662–685.

Chen H, Lim N (2017) How does team composition affect effort in contests? A theoretical and experimental analysis. *J. Mark. Res.* 54(1):44–60.

Chen SF, Dias R, Evans D, Salfati EL, Liu S, Wineinger NE, Torkamani A (2020) Genotype imputation and variability in polygenic risk score estimation. *Genome Med.* 12(1):100.

Choi SW, Mak TSH, O'Reilly PF (2020) Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* 15(9):2759–2772.

Chung DJ, Steenburgh T, Sudhir K (2014) Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans. *Mark. Sci.* 33(2):165–187.

Churchill GA, Ford NM, Hartley SW, Walker OC (1985) The determinants of salesperson performance: A meta-analysis. *J. Mark. Res.* 22(2):103–118.

Collins FS, Morgan M, Patrinos A (2003) The Human Genome Project: Lessons from large-scale biology. *Science* 300(5617):286–290.

Costa PT, McCrae RR (1992) *Revised NEO personality inventory (NEO-PI-R) and Neo five-factor inventory (NEO-FFI)* (Psychological Assessment Resources, Odessa, FL).

Daviet R, Nave G, Wind J (2022) Genetic Data: Potential Uses and Misuses in Marketing. *J. Mark.* 86(1):7–26.

Dudbridge F, Gusnanto A (2008) Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* 32(3):227–234.

Faul F, Erdfelder E, Lang AG, Buchner A (2007) G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39(2):175–191.

Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat. Rev. Genet.* 7(2):85–97.

Fine S, Nevo B (2008) Too smart for their own good? A study of perceived cognitive overqualification in the workforce. *Int. J. Hum. Resour. Manag.* 19(2):346–355.

Ford NM (1983) *Research perspectives on the performance of salespeople: Selected readings* (Marketing Science Inst.).

Ford NM, Walker OC, Churchill GA, Hartley SW (1987) Selecting successful salespeople: A meta-analysis of biographical and psychological selection criteria. *Rev. Mark.* 10:90–131.

Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, et al. (2004) Assessing the impact of population stratification on genetic association studies. *Nat. Genet.* 36(4):388–393.

Friedman NP, Banich MT, Keller MC (2021) Twin studies to GWAS: there and back again. *Trends Cogn. Sci.* 25(10):855–869.
Gilly A, Southam L, Suveges D, Kuchenbaecker K, Moore R, Melloni GE, Hatzikotoulas K, Farmaki AE, Ritchie G, Schwartzentruber J (2019) Very low-depth whole-genome sequencing in complex trait association studies. *Bioinformatics* 35(15):2555–2561.
Given LM (2008) *The Sage encyclopedia of qualitative research methods* (Sage publications).
Grégoire DA, Barr PS, Shepherd DA (2010) Cognitive processes of opportunity recognition: The role of structural alignment. *Organ. Sci.* 21(2):413–431.
Harden KP, Koellinger PD (2020) Using genetics for social science. *Nat. Hum. Behav.* 4(6):1–10.
Holland JL (1997) *Making vocational choices: A theory of vocational personalities and work environments* (Psychological Assessment Resources).
Homburger JR, Neben CL, Mishne G, Zhou AY, Kathiresan S, Khera AV (2019) Low coverage whole genome sequencing enables accurate assessment of common variants and calculation of genome-wide polygenic scores. *Genome Med.* 11(1):1–12.
Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genet.* 5(6):e1000529.
Hu S, Zhang J, Zhu Y (2022) Zero to One: Sales Prospecting with Augmented Recommendation. (January 12) https://papers.ssrn.com/abstract=4006841.
Huang G, Sudhir K (2020) The causal effect of service satisfaction on customer loyalty. *Manag. Sci.* Forthcoming.
Hurtz GM, Donovan JJ (2000) Personality and job performance: The Big Five revisited. *J. Appl. Psychol.* 85(6):869.
Jain S (2012) Self-control and incentives: An analysis of multiperiod quota plans. *Mark. Sci.* 31(5):855–869.
Kalra A, Shi M (2001) Designing optimal sales contests: A theoretical perspective. *Mark. Sci.* 20(2):170–193.
Koellinger PD, Harden KP (2018) Using nature to understand nurture. *Science* 359(6374):386–387.
Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, Benonisdottir S, et al. (2018) The nature of nurture: Effects of parental genotypes. *Science* 359(6374):424–428.
Laffont JJ, Martimort D (2002) *The theory of incentives: the principal-agent model* (Princeton university press, New Jersey).
Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Linnér RK (2018) Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nat. Genet.* 50(8):1112.
Leong SM, Busch PS, John DR (1989) Knowledge bases and salesperson effectiveness: A script-theoretic analysis. *J. Mark. Res.* 26(2):164–178.
Lervåg A, Hulme C (2009) Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychol. Sci.* 20(8):1040–1048.

Lim N, Ahearne MJ, Ham SH (2009) Designing sales contests: Does the prize structure matter? *J. Mark. Res.* 46(3):356–371.
Lo MT, Hinds DA, Tung JY, Franz C, Fan CC, Wang Y, Smeland OB, et al. (2017) Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* 49(1):152–156.
Loveland JM, Lounsbury JW, Park SH, Jackson DW (2015) Are salespeople born or made? Biology, personality, and the career satisfaction of salespeople. *J. Bus. Ind. Mark.* 30(2):233–240.
Madole JW, Harden KP (2021) Building Causal Knowledge in Behavior Genetics. *Behav. Brain Sci.*:1–76.
Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461(7265):747–753.
Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 11(7):499–511.
Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, Derks EM (2018) A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* 27(2):e1608.
Marjoram P, Zubair A, Nuzhdin SV (2014) Post-GWAS: where next? More samples, more SNPs or more biology? *Heredity* 112(1):79–88.
Mayer D, Greenberg HM (1964) What makes a good salesman. *Harv. Bus. Rev.* 42(4):119–125.
Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, Savage JE, et al. (2018) Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* 50(7):920–927.
Nofal AM, Nicolaou N, Symeonidou N, Shane S (2018) Biology and Management: A Review, Critique, and Research Agenda. *J. Manag.* 44(1):7–31.
Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nat. Genet.* 40(5):646–649.
Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, et al. (2016) Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533(7604):539–542.
Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM, Sidorenko J, et al. (2022) Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* 54(4):437–449.
Oren L, Caduri A, Tziner A (2013) Intergenerational occupational transmission: Do offspring walk in the footsteps of mom or dad, or both? *J. Vocat. Behav.* 83(3):551–560.
Papageorge NW, Thom K (2020) Genes, education, and labor market outcomes: evidence from the health and retirement study. *J. Eur. Econ. Assoc.* 18(3):1351–1399.
Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, Gupta N, Neale BM, Daly MJ, Sklar P (2012) Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat. Genet.* 44(6):631–635.
Periatt JA, LeMay SA, Chakrabarty S (2004) The selling orientation–customer orientation

(SOCO) scale: Cross-validation of the revised version. *J. Pers. Sell. Sales Manag.* 24(1):49–54.

Phocas F (2022) Genotyping, the Usefulness of Imputation to Increase SNP Density, and Imputation Methods and Tools. Ahmadi N, Bartholomé J, eds. *Genomic Predict. Complex Traits*. Methods in Molecular Biology. (Springer US, New York, NY), 113–138.

Polderman TJ, Benyamin B, De Leeuw CA, Sullivan PF, Van Bochoven A, Visscher PM, Posthuma D (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* 47(7):702–709.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38(8):904–909.

Purcell SM, Wray R, Stone L, Visscher M, O'Donovan C, Sullivan F, Sklar P, Ruderfer M, McQuillin A, Morris W (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460(7256):748–752.

Rao RS, Viswanathan M, John G, Kishore S (2021) Do Activity-Based Incentive Plans Work? Evidence from a Large-Scale Field Intervention. *J. Mark. Res.* 58(4):686–704.

Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, Westra HJ, Shakhbazov K, Abdellaoui A, Agrawal A (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340(6139):1467–1471.

Ryan L (2017) Ten unmistakable signs you should be in sales. *Forbes*. Retrieved (July 29, 2020), https://www.forbes.com/sites/lizryan/2017/08/04/ten-unmistakable-signs-you-should-be-in-sales/.

Sacerdote B (2007) How large are the effects from changes in family environment? A study of Korean American adoptees. *Q. J. Econ.* 122(1):119–157.

Salgado JF (1997) The Five Factor Model of personality and job performance in the European Community. *J. Appl. Psychol.* 82(1):30.

Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, Strawbridge RJ, Pers TH, Fischer K, Justice AE (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518(7538):187–196.

Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, et al. (2023) The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* 51(D1):D977–D985.

Spiro RL, Weitz BA (1990) Adaptive selling: Conceptualization, measurement, and nomological validity. *J. Mark. Res.* 27(1):61–69.

Sunder S, Kumar V, Goreczny A, Maurer T (2017) Why do salespeople quit? An empirical examination of own and peer effects on salesperson turnover behavior. *J. Mark. Res.* 54(3):381–397.

Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet.* 3(5):391–397.

Tao L, Liang H, Wen B, Huang T (2022) Between nature and nurture: The genetic overlap between psychological attributes and selection into public service employment. *Public Adm.*

Rev.:puar.13582.
Tett RP, Burnett DD (2003) A personality trait-based interactionist model of job performance. *J. Appl. Psychol.* 88(3):500.
Thomas RW, Soutar GN, Ryan MM (2001) The selling orientation-customer orientation (SOCO) scale: A proposed short form. *J. Pers. Sell. Sales Manag.* 21(1):63–69.
Torgeson JK, Wagner RK, Rashotte CA (1999) *Test review: Test of word reading efficiency (TOWRE)* (Pro-ed).
Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D (2021) Genome-wide association studies. *Nat. Rev. Methods Primer* 1(1):59.
Van Leeuwen EM, Kanterakis A, Deelen P, Kattenberg MV, Slagboom PE, De Bakker PI, Wijmenga C, Swertz MA, Boomsma DI, Van Duijn CM (2015) Population-specific genotype imputations using minimac or IMPUTE2. *Nat. Protoc.* 10(9):1285–1296.
Verbeke W, Dietz B, Verwaal E (2011) Drivers of sales performance: a contemporary meta-analysis. Have salespeople become knowledge brokers? *J. Acad. Mark. Sci.* 39(3):407–428.
Verma SS, de Andrade M, Tromp G, Kuivaniemi H, Pugh E, Namjou-Khales B, Mukherjee S, et al. (2014) Imputation and quality control steps for combining multiple genome-wide datasets. *Front. Genet.* 5.
Walker OC, Churchill GA, Ford NM (1977) Motivation and performance in industrial selling: Present knowledge and needed research. *J. Mark. Res.* 14(2):156–168.
Weinfurter DJ (2017) Effective sales talent, the elusive dream. *Forbes*. Retrieved (July 29, 2020), .
Weitz BA, Sujan H, Sujan M (1986) Knowledge, motivation, and adaptive behavior: A framework for improving selling effectiveness. *J. Mark.* 50(4):174–191.
Wies S, Bleier A, Edeling A (2022) EXPRESS: Finding Goldilocks Influencers: How Follower Count Drives Social Media Engagement. *J. Mark.*:00222429221125131.
Wies S, Hoffmann AOI, Aspara J, Pennings JM (2019) Can advertising investments counter the negative impact of shareholder complaints on firm value? *J. Mark.* 83(4):58–80.
Williams C (2017) Sales is not easy - and neither is finding good salespeople. *Forbes*. Retrieved (July 29, 2020), .
Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Kutalik Z, Amin N (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46(11):1173–1186.
Zhang Y, Zhang L, Shu H, Xi J, Wu H, Zhang Yang, Li P (2012) Universality of categorical perception deficit in developmental dyslexia: an investigation of Mandarin Chinese tones. *J Child Psychol Psychiatry*:9.
Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q (2020) A comprehensive survey on transfer learning. *Proc. IEEE* 109(1):43–76.
Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci.* 109(4):1193–1198.