Macroeconomic Dynamics, **26**, 2022, 1833–1890. Printed in the United States of America.
doi:10.1017/S1365100521000092

GENETIC, CULTURAL, AND HISTORICAL DETERMINANTS OF KNOWLEDGE CREATION

Erkan Gören
Carl von Ossietzky University Oldenburg

Knowledge creation has been a pivotal ingredient of endogenous growth theory to understand differences in standards of living across countries. Yet, the identification of key drivers explaining cross-country differences in knowledge creation still remains a topic of central interest in this research field. In this paper, I provide a framework to hypothesize and empirically test the persistent effects of novelty-seeking traits on cross-country differences in scientific knowledge creation. The results suggest a positive and statistically significant relationship between both outcomes that is consistent with the hypothesis that the prevalence of novelty-seeking traits in society facilitates scientific knowledge creation through beneficial human behaviors related to risk-taking and explorative behavior. The empirical findings remain qualitatively unaffected when controlling for additional historical, biogeographical, and socioeconomic factors that appear as additional important determinants in the creation of scientific knowledge in society.

Keywords: Novelty-Seeking Traits, Creative Destruction, Endogenous Growth, Knowledge Creation

1. INTRODUCTION

Models of economic growth that endogenize the pace of technological progress suggest that vertical and horizontal product innovations contribute significantly to the level of economic development [Romer (1990); Grossman and Helpman (1991); Aghion and Howitt (1992)]. A central result is that differences in income per capita across countries are attributable to differences in the level of technological development. In these models, the rate of technological progress is determined by the knowledge stock in society overall.

Several policy implications have been drawn from endogenous growth models to facilitate knowledge creation in the aggregate economy such as the need

I thank the Editor, William A. Barnett, an Associate Editor, and a anonymous Referee for their helpful comments and suggestions that substantially improved the paper. I would also like to thank Jürgen Bitzer, Andreas Lichter, and the session participants at the Annual Conference of the Verein für Socialpolitik (VfS) 2018, University of Freiburg (im Breisgau) for useful comments and suggestions. All remaining errors are my own. Address correspondence to: Erkan Gören, Carl von Ossietzky University Oldenburg, School of Computing Science, Business Administration, Economics, and Law (Faculty II), Institute of Economics, Building A5, 26111 Oldenburg, Germany. e-mail: erkan.goeren@uni-oldenburg.de. Phone: +49 441 798 4292.

 1469-8056/21

to protect the intellectual property rights of innovators, to allocate subsidies to research and development activities, and to make investments in human capital. Although these factors appear important for the accumulation of knowledge in society, they refer to proximate causes of economic development and do not reveal in greater detail why countries differ in these observed characteristics.

In this paper, I provide positive evidence that differences in knowledge creation across countries are to some extent determined by deep-rooted historical factors. Specifically, I show that the prevalence of novelty-seeking traits—measured by the frequency of the 2- and 7-repeat allele variants of the human dopamine D4 receptor (DRD4) gene—has a positive relationship with scientific research output—as a proxy for knowledge creation—in a large cross section of countries. The argument is that the type of behavioral attributes related to novelty-seeking individuals such as risk-taking, creative, and explorative activity fosters the accumulation of knowledge in the aggregate economy. Therefore, to the best of the author's knowledge, this is the first study in economics to provide evidence that the kind of behavioral attributes frequently linked to the Schumpeterian notion of *creative destruction* stimulates technological progress through increased knowledge creation in a cross-national context.

The underlying theory builds on the Schumpeterian-inspired endogenous growth model of Aghion and Howitt (1992). I extend the basic model framework to provide a micro-foundation of the innovation process that describes the occurrence of innovations in the research sector as the outcome of two individual random processes regarding the probability of success and the amount of research projects channeled to the research sector. The model predicts that the rate of innovations that occur in the economy is, among others, a positive function of the prevalence of novelty-seeking traits in society. This finding is consistent with the notion that innovation activity is stimulated through beneficial economic attitudes related to risk-taking, creativity, explorative, and entrepreneurial behavior.

I combine various data sources to investigate the relationship between knowledge creation and the prevalence of novelty-seeking traits in society. To measure the extent of knowledge creation in society, I use the number of scientific and technical journal articles during the period 1981–2016 that have been published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. It is generally acknowledged in the relevant literature that a country's scientific knowledge base plays an important role in the process of innovation by private firms.[1]

One of the main challenges in the empirical analysis pertains to the operationalization and/or measurement of a latent and complex personality trait such as novelty-seeking behavior that includes a variety of possible behavioral dimensions (e.g., explorative, creative, and risk-taking activity). I follow the suggestion of Faraone et al. (2014) and use a set of genetic biomarkers that are physiologically related to the development of specific personality traits. In particular, novelty-seeking-related behavioral outcomes are natural candidates to

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

be biologically influenced by dopamine genes. The argument is that the neurophysiological mechanisms that occur in the dopamine system are of pivotal importance for the degree of functionality relevant for the control of locomotion, reward, cognition, and emotional stability [Oak et al. (2000)]. Consistent with this hypothesis, two influential candidate gene association studies reported a significant association between the DRD4 exon III gene and self-reported measures of novelty-seeking test score ratings [Benjamin et al. (1996); Ebstein et al. (1996)].

Based on the initial evidence from molecular genetics research, I use the frequency of the 2- and 7-repeat allele variants of the DRD4 exon III gene that population geneticists have found to be physiologically related to the development of a complex human phenotype such as novelty-seeking behavior. This novel measure has been constructed in Gören (2017) by matching the distribution of ethnic groups in the Alesina et al. (2003) ethnicity data to the population genome data of the human dopamine D4 receptor gene in Gören (2016) using information on the classification of ethno-linguistic groups from the *Ethnologue* database [Global Mapping International (2010)]. The narrative of this approach is that a higher frequency of these allelic variants found among populations helps to predict the tendency of people to display novelty-seeking-related behavioral outcomes in society.

The baseline results indicate a positive and statistically significant association between the number of scientific and technical journal publications *per capita* and the country-level DRD4 exon III 2- and 7-repeat allele frequency measure. This result is robust to the inclusion of microgeographic, land productivity, climatic, health, legal, religious, historical, and regional factors. The control variables in this list constitute alternative determinants of scientific knowledge creation that might be additionally correlated with the prevalence of novelty-seeking traits in society. Specifically, the inclusion of biogeographic controls rules out concerns that the effect of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure on scientific knowledge creation simply captures the issue of geographic proximity. This possibility is of considerable importance since previous research has suggested that the between-population distribution of DRD4 exon III allele variants is the result of natural selection that has been ongoing since the exodus of the human species out of East Africa [Gören (2016)].

The baseline estimate suggests that an increase in the country-level DRD4 exon III 2- and 7-repeat allele frequency value of Poland (i.e. 0.1992) to the level of the United States (i.e. 0.2636) would, ceteris paribus, result in a increase in the number of scientific and technical publications *per capita* by about 25.16%. This increase in the dependent variable is equivalent to moving Poland from the 77.34th percentile to the 78.91st percentile in the distribution of the number of scientific and technical journal publications *per capita*.

I conduct a series of additional sensitivity tests to examine the robustness of the main empirical findings to various model specifications. First of all, I show that the main results are robust to the issue of genetic, ethnic, linguistic, and religious diversity. This observation effectively rules out the possibility that the estimated relationship between scientific knowledge creation and the country-level DRD4 exon III 2- and 7-repeat allele frequency measure merely reflects unobserved factors related to various aspects of societal diversity. The results are also robust to the inclusion of various technological distance controls that capture the notion of human barriers to the diffusion of knowledge across countries (i.e. genetic, linguistic, and religious distance from the technological frontier). Moreover, I provide regression coefficients that account for the country's technological and human capability factors, which appear to be of considerable importance for the creation of scientific knowledge in society. Specifically, the baseline estimates are not affected by the inclusion of factors such as national research and development efforts, infrastructure quality, and a human capital index.

Finally, I use a set of economic preference factors to rule out concerns about alternative interpretations regarding the pleiotropic effects of the DRD4 exon III polymorphism on other personality outcomes besides novelty-seeking behavior, such as patience, willingness to take risks, and various social preference factors (i.e. altruism, negative and positive reciprocity, and trust). Reassuringly, the main results remain qualitatively unaffected by the inclusion of these economic preference controls.

This paper contributes to the literature on some deep-rooted factors of cross-country differences in standards of living. The importance of novelty-seeking traits for the rate of technological progress has been analyzed theoretically by Galor and Michalopoulos (2012). The authors argue that the kind of human behaviors frequently ascribed to novelty-seeking individuals (e.g., entrepreneurial activity and risk-taking) is the main drivers of innovation activity in the process of economic development. The study by Gören (2017) provides the first evidence of a statistically significant, inverted U-shaped association between the prevalence of novelty-seeking traits in society and comparative economic development. This observation is suggestive of the potential "benefits" (e.g., knowledge creation) and "costs" (e.g., educational and occupational disadvantages) experienced by novelty-seeking individuals in the aggregate economy. Other studies have documented the role of additional specific genetic markers in pre-colonial economic development, contemporary life expectancy, cultural attitudes, and innovations [Cook (2014, 2015); Gorodnichenko and Roland (2017)]. In addition, Ashraf and Galor (2013) examine the non-monotonic influence of overall human genetic diversity on pre-colonial and contemporary differences in income per capita in a cross section of countries. Spolaore and Wacziarg (2009) and (2012) found that a country's relative genetic distance from the technological frontier contributed significantly to differences in economic development and to the diffusion of technological innovations across borders. This finding is in line with the notion that genetic distance captures deep-rooted historical and cultural differences among populations that hinder the diffusion of technological and institutional improvements across countries.

The paper is organized as follows. Section 2 presents the economic argument to highlight the importance of novelty-seeking traits in the process of scientific knowledge creation and, thus, economic development. Section 3 presents the econometric framework used in the empirical analysis of the relationship between scientific knowledge creation and the county-level DRD4 exon III 2- and 7-repeat allele frequency measure. Section 4 provides a detailed description of the main variables employed in the empirical analysis. The main empirical results are discussed in Section 5. Additional sensitivity tests are reported in Section 6. Finally, Section 7 concludes by summarizing the main findings.

2. A SCHUMPETERIAN GROWTH MODEL OF KNOWLEDGE CREATION

In this section, I develop an endogenous economic growth model to conceptualize the main hypothesis proposed in this paper regarding the beneficial effects of novelty-seeking individuals in the process of knowledge creation in the aggregate economy. The underlying theoretical framework is the Schumpeterian-inspired model of economic growth proposed in Aghion and Howitt (1992)'s study, where growth is generated through repeated improvements in the quality of intermediate goods. This model framework is particularly convenient for studying the implications of novelty-seeking individuals in the process of economic growth. The reason is that the type of attributes related to the Schumpeterian notion of 'creative destruction'—risk-taking, creativity, and entrepreneurship—is those frequently attributed to the kind of human behaviors usually found in novelty-seeking individuals.

The underlying source of technological progress in this model of economic growth is the research sector, where the flow of workers channeled to R&D activities influences the rate of innovations that occur in the economy. It is worth mentioning that in this model framework, the arrival rate per researcher, that is, the rate at which innovations occur randomly in the aggregate economy, is exogenously given and that its main determinants remain largely unexplained. However, this assumption might be misleading given that research productivity differs substantially across countries. Specifically, Figure 1 demonstrates the notable differences that exist in contemporary research productivity across countries, as indicated by the number of scientific and technical publications per 1000 people in the years 1981–2016. The corresponding summary statistics are reported in Table 1. The mean value of scientific and technical publications per 1000 people is about 0.2058 with a relatively large standard deviation of 0.3558 across the available set of 171 countries.

Given the fact that research productivity differs substantially across countries, I extend the basic model framework in Aghion and Howitt (1992) by providing a microeconomic-based derivation of the source of technological progress in the aggregate economy. To accomplish this task, I define innovation as the outcome of two individual random processes.

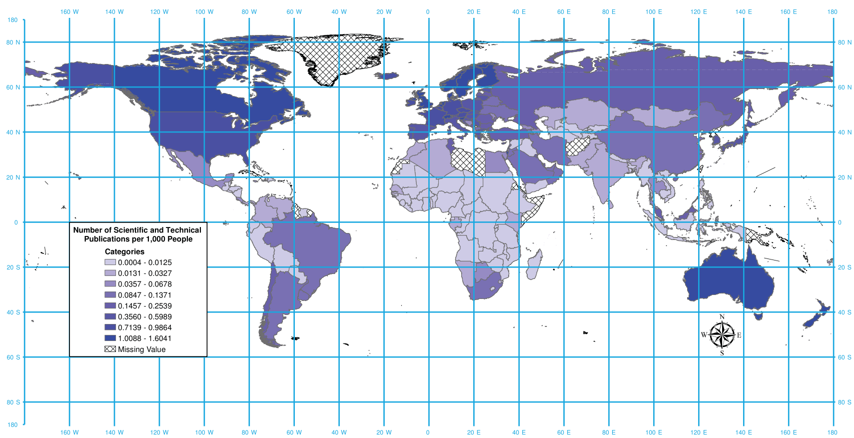
In the first process, the probability of successful innovations occurring out of a sequence of randomly distributed research projects has a binomial distribution.

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

TABLE 1. Number of scientific and technical publications per 1000 people across continents

Variable	N	Mean	SD	Minimum	Maximum
Scientific Publications per 1000 People	171	0.2058	0.3558	0.0004	1.6041
Africa	50	0.0152	0.0271	0.0004	0.1496
Americas	33	0.1000	0.2521	0.0010	1.1384
Asia	45	0.1238	0.2542	0.0006	1.3117
Europe	40	0.5817	0.4167	0.0239	1.6041
Oceania	3	0.7629	0.6182	0.0550	1.1966

Notes: This table shows basic summary statistics for the number of scientific and technical publications per 1000 people in the years 1981–2016 across continents. It refers to scientific publications that have been published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. See the main text for additional details regarding data construction and sources.



Notes: This map shows the worldwide distribution of the number of scientific and technical publications per 1000 people in the years 1981– 2016 across countries. It refers to scientific publications that have been published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. See the main text for additional details regarding data construction and sources.

FIGURE 1. Number of scientific and technical publications per 1000 people across countries.

I endogenize the corresponding probability of success as a function of the extent of novelty-seeking traits in society. The second source of uncertainty refers to the number of research projects that occur randomly within the unit time interval. This circumstance is modeled according to a Poisson distribution with an endogenous arrival rate, which is a function of the country's technological and human resource base. Then, the innovation process in the aggregate economy is

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

fully determined by the corresponding individual random processes, which follows a Poisson distribution, with a corresponding arrival rate that is a function of the prevalence of novelty-seeking traits in society. The theoretical model suggests that a higher prevalence of novelty-seeking traits in society, ceteris paribus, increases the extent of knowledge creation through a higher propensity of research projects toward success. This in turn stimulates higher long-run economic growth in the aggregate economy. In the following, I provide an in-depth discussion of the proposed theoretical framework. This sets the basis for the econometric specification to estimate the impact of the prevalence of novelty-seeking traits on scientific knowledge creation.

2.1. Theoretical Framework

Final goods sector. The economy produces a final product y using a single intermediate good m subject to the following Cobb–Douglas production technology,

$$y_t = A_t m_t^{\alpha}, \tag{1}$$

where y is the output, A is the level of technology, m is the amount of intermediate goods, and $0 < \alpha < 1$ indicates the output elasticity with respect to intermediate goods. For the sake of simplicity, I leave out population growth, human, and physical capital accumulation in the process of economic growth to keep the analysis as simple as possible. As usual in Schumpeterian growth models, the subscript $t = \{0, 1, 2, \ldots\}$ does not refer to continuous time but rather to the start of the t-th innovation in the aggregate economy. The main idea behind this model framework is that the economy still uses the highest quality of intermediate goods in the production process until it is rendered obsolete by the invention of a new technology, that is, $A_{t+1} > A_t$.

Intermediate goods sector. Each individual is endowed with one unit of labor that they can allocate freely between the intermediate goods and research sector. The intermediate good is produced with a linear technology that employs labor as the only production factor,

$$m_t = \ell_t^M, \tag{2}$$

where ℓ_t^M denotes the flow of labor in the intermediate goods sector of the t-th innovating firm. It is assumed that the intermediate goods sector is characterized by monopolistic competition. An innovating firm gets a temporary patent to earn profits in the intermediate goods sector until it is replaced by outside firms that conduct research. The duration of the patent is indefinite due to uncertainty regarding the discovery of new inventions in the research sector.

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

Research sector. Growth in this model is generated from successful innovations in the research sector, which improves the quality of intermediate goods up to a productivity factor $\gamma > 1$ according to

$$A_t = A_0 \gamma^t, \tag{3}$$

where A_0 refers to the initial level of technology, which might be a function of historical and biogeographic factors. In the following, I provide a microeconomic-based derivation of the innovation process that models the research output as the result of two individual random processes. This model framework proves useful to highlight the importance of the prevalence of novelty-seeking traits for the process of innovation activity and to disentangle its impact from other historical, socioeconomic, and technological factors.

Indicate with τ the continuous time variable which is of unity interval (i.e. $\Delta\tau = 1$). Denote with T_i the i-th research project in the aggregate economy which can take two possible states. These two states refer to the outcome of the innovation process, which might result in a successful new invention ($T_i = 1$) or might not ($T_i = 0$). In probability theory, the Bernoulli distribution is used quite frequently to model random processes that take two possible outcomes. According to this distribution, the probability mass function of the random variable T_i with outcomes $k = \{0, 1\}$ is

$$f_{T_i}(T_i = k \mid p) = \begin{cases} p & \text{if} \quad k = 1 \quad \text{(successful innovation)}, \\ 1 - p & \text{if} \quad k = 0 \quad \text{(unsuccessful innovation)}, \end{cases} \tag{4}$$

where $p \in [0, 1]$ refers to the probability of success if the i-th research project results in a new invention (i.e. $T_i = 1$) and $1 - p$ if not (i.e. $T_i = 0$).

A main finding from endogenous growth models is that differences in income per capita across countries are attributable to differences in the level of technology, as indicated in equation (3). The global pattern of technological performance suggests that countries differ substantially in levels of innovation activity. A possible approach to modeling such technological differences across countries is to endogenize the probability of success in the Bernoulli distribution that describes the outcome of individual research projects. Specifically, it is desirable to assume that the success probability of research projects is a function of the prevalence of novelty-seeking traits in society. The argument is that novelty-seeking individuals are quite successful in the creation of new knowledge or innovations due to their innate explorative, risk-taking, creative, and entrepreneurial behavior, which pushes the technological frontier forward [Galor and Michalopoulos (2012); Gören (2017)]. From an evolutionary perspective, it has been argued that novelty-seeking traits are quite beneficial in risky and time-critical environments [Jensen et al. (1997); Gören (2016)].

Thus, the probability of successful innovations is modeled according to $p = p(z, \nu)$, where ν is the prevalence of novelty-seeking traits in society. The parameter z refers to additional historical, biogeographical, technological, and human capability factors that characterize the economic environment, where individuals engage in cooperative social behavior. The specific mechanisms by which these factors influence the innovation process are discussed in greater detail below. However, it is worth mentioning that the narrative of this approach is related to the basic idea that historical, geographical, and socioeconomic circumstances might act as barriers to the diffusion of technological improvements across borders that in turn negatively affect a country's own innovation process [Basu and Weil (1998)].

ASSUMPTION 1. *The function* $p : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ *is twice differentiable, monotonically increasing, and of diminishing marginal returns with respect to* z *and* ν*, and satisfies the following conditions:*

$$p_z(z,\nu) \equiv \frac{\partial p(z,\nu)}{\partial z} > 0, \qquad p_\nu(z,\nu) \equiv \frac{\partial p(z,\nu)}{\partial \nu} > 0, \tag{5}$$

$$p_{zz}(z,\nu) \equiv \frac{\partial^2 p(z,\nu)}{\partial z^2} \leq 0, \qquad p_{\nu\nu}(z,\nu) \equiv \frac{\partial^2 p(z,\nu)}{\partial \nu^2} \leq 0, \tag{6}$$

$$p(0,\nu) = 0, \qquad p(z,0) = 0, \tag{7}$$

for all $z \in [0, 1]$*, and* $\nu \in [0, 1]$.

In addition, the cross-partial derivatives $p_{\varphi\varphi'}(z,\nu) \equiv \partial^2 p(z,\nu)/\partial\varphi\partial\varphi' > 0$, where $\varphi, \varphi' = \{z, \nu\}$, have a positive sign, indicating possible complementarities between the prevalence of novelty-seeking traits and socioeconomic factors. The idea is that the benefits of novelty-seeking traits for the creation of knowledge in society might be dependent on prevailing institutional and economic circumstances that make it easier to reap gains from explorative behavior. For example, novelty-seeking individuals share many behavioral symptoms (e.g., impulsive, risk-taking, and sensation-seeking behavior) usually found among individuals diagnosed with ADHD, resulting in educational and occupational difficulties in modern societies that are characterized by clear social hierarchies [Dannemann and Gören (2018)]. Thus, countries that are unable to develop effective educational and labor market strategies to mitigate the potential costs of novelty-seeking traits would find it hard to benefit from their explorative nature for the aggregate economy [Gören (2017)].

Usually, economies are engaged in more than one research project per unit time interval. A sequence $T = \sum_{i=1}^{N} T_i$ of N identically independently distributed ($i.i.d.$) random research projects T_i with identical individual success probabilities $p(z,\nu)$ has a binomial distribution with parameters $N \in \mathbb{N}$ and $p(z,\nu) \in [0, 1]$ with a corresponding probability mass function,

$$f_T\left[T = t \mid N, p(z,\nu)\right] = \binom{N}{t} p(z,\nu)^t \left[1 - p(z,\nu)\right]^{(N-t)}, \tag{8}$$

where $t = \{0, 1, 2, \ldots, N\}$ is the number of successful innovations during the unit time interval. This formulation illustrates the uncertainty of the innovation process, indicating that of N random research projects, one could expect on average

$\mathbb{E}\left[\sum_{i=1}^{N} T_i\right] = N \times p(z,\nu)$ successful innovations with variance $\mathbb{V}\left[\sum_{i=1}^{N} T_i\right] = N \times p(z,\nu)\left[1-p(z,\nu)\right]$.

Next, consider another source of technological heterogeneity: It is conceivable that N, indicating the number of research projects in the unit time interval, differs across countries too. In particular, differences in the flow of researchers (ℓ_t^R), the amount of government resources channeled to R&D activities (R), and the economic environment (ψ) might explain widespread variations in the extent of research activity N across countries. In this paper, I model the number of research projects per unit time interval using a Poisson distribution. In probability theory, the Poisson distribution expresses the probability of randomly occurring events (here: number of research projects N) in a fixed time interval (here: $\Delta\tau = 1$) according to the following probability density function:

$$f_N\left[N=n \mid \mu\left(\psi,\ell_t^R,R\right)\right] = \frac{\left[\mu\left(\psi,\ell_t^R,R\right)\right]^n}{n!} e^{-\mu\left(\psi,\ell_t^R,R\right)}, \tag{9}$$

with a corresponding arrival rate $\mu\left(\psi,\ell_t^R,R\right) \equiv (1-\psi)\phi(\ell_t^R,R) \in \mathbb{R}_{>0}$. The arrival rate indicates the expected number of research projects per unit time interval, that is, $\mathbb{E}[N] = (1-\psi)\phi(\ell_t^R,R)$ with variance $\mathbb{V}[N] = (1-\psi)\phi(\ell_t^R,R)$, which is identical to its expected value. The constant parameter $\psi \in [0,1]$ captures potential market distortions in the research sector (e.g., widespread corruption, rent-seeking activities, or growth-distorting government taxes) that undermine its productivity.

ASSUMPTION 2. *The economy generates research projects according to the research production technology $\phi : \mathbb{R}_+^2 \to \mathbb{R}_+$ which is subject to constant returns to scale and satisfies the following conditions:*

$$\phi_{\ell_t^R}(\ell_t^R,R) \equiv \frac{\partial \phi(\ell_t^R,R)}{\partial \ell_t^R} > 0, \qquad \phi_R(\ell_t^R,R) \equiv \frac{\partial \phi(\ell_t^R,R)}{\partial R} > 0, \tag{10}$$

$$\phi_{\ell_t^R\ell_t^R}(\ell_t^R,R) \equiv \frac{\partial^2 \phi(\ell_t^R,R)}{\partial \left(\ell_t^R\right)^2} \le 0, \qquad \phi_{RR}(\ell_t^R,R) \equiv \frac{\partial^2 \phi(\ell_t^R,R)}{\partial R^2} \le 0, \tag{11}$$

$$\phi(0,R) = 0, \qquad \phi(\ell_t^R,0) = 0, \tag{12}$$

for all $\ell_t^R \ge 0$, and $R \ge 0$.

The innovation process in the research sector is then determined by the corresponding probability functions of the individual random processes, as described in equations (8) and (9), according to the following composite probability density function:

$$f_T\left[T=t \mid \lambda\left(\psi,\ell_t^R,R,z,\nu\right)\right] = \sum_{n=0}^{\infty} f_T\left[T=t \mid N=n, p(z,\nu)\right] \times f_N\left[N=n \mid \mu\left(\psi,\ell_t^R,R\right)\right]. \tag{13}$$

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

This expression corresponds to the mixture of two individual probability distributions to a weighted probability distribution model for the occurrence in the number of *successful innovations* $t = \{0, 1, 2, \ldots\}$ per unit time interval in the research sector. In particular, the binomial distribution in equation (8) with the randomly distributed parameter N is weighted over all possible realizations $n = \{0, 1, 2, \ldots\}$ according to the probability density function of the Poisson distribution in equation (9) with arrival rate $\mu\left(\psi, \ell_t^R, R\right)$. The closed-form solution of equation (13) is given by

$$f_T\left[T = t \mid \lambda\left(\psi, \ell_t^R, R, z, \nu\right)\right] = \frac{\left[\lambda\left(\psi, \ell_t^R, R, z, \nu\right)\right]^t}{t!} e^{-\lambda\left(\psi, \ell_t^R, R, z, \nu\right)}, \qquad \textbf{(14)}$$

where $\lambda(\psi, \ell_t^R, R, z, \nu) \equiv (1-\psi)\phi(\ell_t^R, R)p(z, \nu) \in \mathbb{R}_{>0}$.

In summary, the discretely occurring events that comprise the number of successful innovations per unit time interval follow a Poisson distribution with arrival rate $\lambda(\psi, \ell_t^R, R, z, \nu)$ which, among other things, is a positive function of the prevalence of novelty-seeking traits ν in society. However, the steady-state arrival rate of the number of innovations depends on the flow of labor ℓ_t^R channeled into the research sector. This allocation process is governed by the corresponding profit considerations of economic agents in the intermediate goods and research sector, which I discuss in more detail below.

2.2. The Monopolist's Profit Maximization Problem

The intermediate goods sector is characterized by monopolistic competition. Intermediate goods are produced according to the linear technology $m_t = \ell_t^M$. The monopolist firm stays in the intermediate goods market until it is replaced by the next innovating firm. Since innovations in the economy per unit time interval occur randomly according to a Poisson distribution with arrival rate $\lambda(\psi, \ell_t^R, R, z, \nu)$, the length of time between two consecutive innovations, that is, $\Delta(t+1, t) = \tau_{t+1} - \tau_t$, is exponentially distributed with parameter $\lambda(\psi, \ell_t^R, R, z, \nu)$, where the expected length of time between two innovations is inversely related to the Poisson arrival rate, that is, $\mathbb{E}[\Delta(t+1, t) = \tau_{t+1} - \tau_t] = 1/[\lambda(\psi, \ell_t^R, R, z, \nu)]$ with variance $\mathbb{V}[\Delta(t+1, t) = \tau_{t+1} - \tau_t] = 1/[\lambda(\psi, \ell_t^R, R, z, \nu)]^2$. Thus, the length of the interval in which the monopolist firm can earn profits or remain in the intermediate goods market is negatively related to the arrival rate of innovations in the research sector.

The monopolist firm is faced with the inverse demand function in the final goods sector $p_t = \alpha A_t m_t^{\alpha-1}$ and takes as given the wage rate w_t of skilled labor. It therefore maximizes its flow of profits according to the following profit function:

$$\max_{m_t} \pi_t = \left[p_t m_t - w_t m_t\right], \qquad \textbf{(15)}$$

where π_t refers to profits of the t-th innovating firm. Substituting the equation for the inverse demand function of the intermediate goods in the final goods sector into the monopolist profit function and maximizing with respect to m_t yields the monopolist firm's labor demand function in the intermediate goods sector,

$$\ell_t^M \equiv \widetilde{\ell}^M(\omega_t) = \left(\frac{\alpha^2}{\omega_t}\right)^{\frac{1}{(1-\alpha)}} \quad \Leftrightarrow \quad \omega_t \equiv \widetilde{\omega}(\ell_t^M) = \frac{\alpha^2}{\left(\ell_t^M\right)^{(1-\alpha)}}, \tag{16}$$

where $\omega_t = w_t/A_t$ is the productivity-adjusted wage rate. Thus, the monopolist firm's demand for skilled labor in the intermediate goods sector is inversely related to the productivity-adjusted wage rate. Given equation (16), the monopolist profit flow is given by

$$\pi_t = A_t\widetilde{\pi}(\omega_t), \tag{17}$$

where $\widetilde{\pi}(\omega_t) = (\frac{1-\alpha}{\alpha})(\frac{\alpha^2}{\omega_t^\alpha})^{\frac{1}{(1-\alpha)}}$. This equation expresses the monopolist's profits as a negative function of the productivity-adjusted wage rate ω_t and as a positive function of the level of technology A_t.

2.3. Perfect Competition in the Research Sector

Research is conducted by outside firms that employ skilled labor in a patent-race competitive framework to earn the expected present value of future profit flows in the intermediate goods sector of the next $(t+1)$-th innovation, denoted by V_{t+1}, according to the following expression:

$$V_{t+1} = \int_0^\infty \widetilde{V}_{t+1}(\tau) f_\Delta \left[\Delta\,(t+2, t+1) = \tau \middle| \lambda\left(\psi, \ell_{t+1}^R, R, z, \nu\right)\right] d\tau, \tag{18}$$

where $f_\Delta[\Delta(t+2, t+1) = \tau|\lambda(\psi, \ell_{t+1}^R, R, z, \nu)] = \lambda(\psi, \ell_{t+1}^R, R, z, \nu)$ $e^{-\lambda(\psi,\ell_{t+1}^R,R,z,\nu)\tau}$ is the probability density function of the length of time $\Delta(t+2, t+1) = \tau_{t+2} - \tau_{t+1} \equiv \tau$ between two consecutive innovations, which has an exponential distribution with parameter $\lambda(\psi, \ell_{t+1}^R, R, z, \nu) \in \mathbb{R}_{>0}$. It is worth mentioning that the $(t+1)$-th innovating firm will face uncertainty over its future profit flow because of research conducted by outside firms during the $(t+1)$-th innovation interval. The present value of the future profit flow of the $(t+1)$-th innovating firm until it is replaced by the next $(t+2)$-th innovator, denoted by $\widetilde{V}_{t+1}(\tau)$, is given by

$$\widetilde{V}_{t+1}(\tau) = \int_{\tau_{t+1}}^{\tau_{t+2}} \pi_{t+1} e^{-r(s-\tau_{t+1})} ds. \tag{19}$$

The present value of the future monopolist profit flow $\widetilde{V}$ is a random variable because of the uncertainty regarding the length of time between two consecutive innovations, and r is the constant real interest rate. Thus, the solution to equation (19) is given by

$$\widetilde{V}_{t+1}(\tau) = \frac{\pi_{t+1}\left[1 - e^{-r\tau}\right]}{r}. \tag{20}$$

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

If $\tau = (\tau_{t+2} - \tau_{t+1})$, the length of time of the $(t+1)$-th innovation, approaches infinity in the limit, then the monopolist expected future profit flow would correspond to its value of perpetuity, that is, π_{t+1}/r for $\tau \to \infty$.

Substituting this expression for $\widetilde{V}_{t+1}(\tau)$ in equation (18) and solving the integral with respect to τ gives the following solution for the monopolist firm's expected present value of future profits in the intermediate goods sector,

$$V_{t+1} = \frac{\pi_{t+1}}{r + \lambda\left(\psi, \ell_{t+1}^R, R, z, \nu\right)}, \tag{21}$$

which is similar to the expression in Aghion and Howitt (1992)'s study except for the fact that the Poisson arrival rate $\lambda(\psi, \ell_{t+1}^R, R, z, \nu)$ of successful innovations in this model setup has been derived from two individual random processes in the research sector.

After having defined the value that the outside firm can earn in the intermediate goods sector when it produces a successful innovation, we are now in the position to describe the maximization problem of the research-conducting firm that determines its research employment in the innovation process. Without loss of generality, I assume an infinitesimally small time interval $d\tau$ during the t-th and $(t+1)$-th innovations. During this time interval, the probability that the research-conducting outside firm will produce an innovation allowing it to take over the entire intermediate goods sector is approximately $\lambda(\psi, \ell_t^R, R, z, \nu)d\tau$.[2] Then, the firm's expected profit from conducting research equals $\lambda(\psi, \ell_t^R, R, z, \nu)d\tau V_{t+1}$ with corresponding costs due to research employment in the amount of $w_t d\tau \ell_t^R$ during the small time interval $d\tau$. Hence, the research-conducting outside firm maximizes the following profit function with respect to research employment ℓ_t^R:

$$\max_{\ell_t^R} \left[\lambda\left(\psi, \ell_t^R, R, z, \nu\right) d\tau\, V_{t+1} - w_t d\tau \ell_t^R\right]. \tag{22}$$

The solution to this maximization problem yields the famous research arbitrage equation in the Aghion and Howitt (1992) Schumpeterian growth model,

$$w_t = \lambda_{\ell_t^R}\left(\psi, \ell_t^R, R, z, \nu\right) V_{t+1}, \tag{23}$$

where $\lambda_{\ell_t^R}(\psi, \ell_t^R, R, z, \nu) \equiv \partial\lambda(\psi, \ell_t^R, R, z, \nu)/\partial\ell_t^R$ refers to the first partial derivative of the Poisson arrival rate with respect to research employment. The research arbitrage equation states that outside firms will employ skilled labor in the research sector up to the point where their marginal cost of research w_t equals their expected marginal benefit of research $\lambda_{\ell_t^R}(\psi, \ell_t^R, R, z, \nu)V_{t+1}$.

2.4. Equilibrium Dynamics

In the following, I derive the dynamical system that characterizes the fundamental law of motion for the equilibrium path in a decentralized economy (DE). The research arbitrage equation shown in equation (23) together with the labor market clearing condition $L = \ell_t^R + \widetilde{\ell}^M(\omega_t)$ entirely describes the equilibrium condition

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

in a DE. For the sake of simplicity, I assume that labor supply is identical to population size. In equilibrium, skilled workers in the intermediate goods and research sector must be paid the same wage rate, which is determined by the monopolist inverse labor demand function $\widetilde{\omega}(\ell_t^M)$ in equation (16). Using equations (16), (20), (21), the labor market clearing condition $\ell_t^M = L - \ell_t^R$, and the fact that $A_{t+1} = \gamma A_t$, it follows for the research arbitrage equation after some basic rearrangements that

$$\frac{\widetilde{\omega}(L-\ell_t^R)}{\lambda_{\ell_t^R}\left(\psi, \ell_t^R, R, z, \nu\right)} \geq \frac{\gamma\widetilde{\pi}\left[\widetilde{\omega}(L-\ell_{t+1}^R)\right]}{r+\lambda\left(\psi, \ell_{t+1}^R, R, z, \nu\right)}, \quad \ell_t^R \geq 0, \tag{24}$$

with at least one equality. Following the definition by Aghion and Howitt (1992), $c(\ell_t^R) = \frac{\widetilde{\omega}(L-\ell_t^R)}{\lambda_{\ell_t^R}(\psi,\ell_t^R,R,z,\nu)}$ refers to the "marginal cost of research", and $b(\ell_{t+1}^R) = \frac{\gamma\widetilde{\pi}[\widetilde{\omega}(L-\ell_{t+1}^R)]}{r+\lambda(\psi,\ell_{t+1}^R,R,z,\nu)}$ is the "marginal benefit of research". Then, from equation (16), and Assumptions 1 and 2, it follows that $c(\ell_t^R)$ is strictly increasing and $b(\ell_{t+1}^R)$ strictly decreasing in research employment, and $c(\ell_t^R) \to \infty$ as $\ell_t^R \to L$ and $b(\ell_{t+1}^R) = 0$ if $\ell_{t+1}^R = L$. The equilibrium condition in (24) defines research employment during the current t-th innovation period as a function of research employment in the $(t+1)$-th innovation period in the future. The classic equilibrium dynamics analysis in the standard Aghion and Howitt (1992) Schumpeterian growth model leads to the following central proposition:

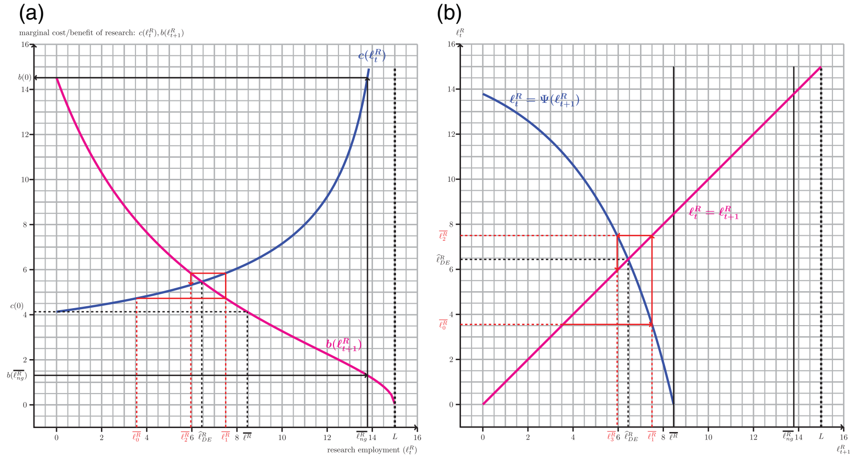
PROPOSITION 1. *There exists a strictly decreasing function* $\Psi \in [0, \overline{\ell^R}]$ *such that current research employment is a negative function of future research:*

$$\ell_t^R = \Psi\left(\ell_{t+1}^R\right). \tag{25}$$

The function $\Psi(\ell_{t+1}^R)$ *is well defined on* $[0, L]$ *and is positive and decreasing if* $c(0) < b(0)$, *where* $c(0) = \frac{\widetilde{\omega}(L)}{\lambda_{\ell_t^R}(\psi,\ell_t^R,R,z,\nu)}|_{\ell_t^R=0}$ *and* $b(0) = \frac{\gamma\widetilde{\pi}[\widetilde{\omega}(L)]}{r}$. *In this case, the critical value* $\overline{\ell^R} < L$ *satisfies the condition* $c(0) = b(\overline{\ell^R})$. *However, if* $\lim_{\ell_{t+1}^R \to L} b(\ell_{t+1}^R) > c(0)$, *then* $\overline{\ell^R} = L$. *In the case where* $c(0) \geq b(0)$, $\Psi(\ell_{t+1}^R) = 0$ *for all* $\ell_{t+1}^R \geq 0$.

The reason for the negative dependency between current and future research employment is that firms are discouraged from investing more in research today by an anticipated decrease in the flow of future profits due to a higher productivity-adjusted wage rate and a shorter length of time between two consecutive innovations, which is triggered by an increase in the Poisson arrival rate of future innovations [Aghion and Howitt (1992)].

To illustrate the model's equilibrium dynamics with respect to the various parameters involved in overall innovation activity, I assume for the sake of simplicity that the probability of successful innovations can be modeled as follows:



The relationship between future and current research

Depicting the functional relationship $\ell_t^R = \Psi\left(\ell_{t+1}^R\right)$

FIGURE 2. Equilibrium dynamics in the Schumpeterian growth economy.

ASSUMPTION 3. *The probability of successful innovations can be modeled according to* $p(z,\nu) = \int_{s=0}^{z}\int_{u=0}^{\nu} h(s)^{\beta} q(u)^{1-\beta} dsdu$, *where* $q(u)$ *is the continuum of certain personality traits (e.g., risk-taking, sensation-seeking, or curiosity)* $u \in [0,\nu]$ *associated with the prevalence of novelty-seeking traits in society. In a similar vein,* $h(s)$ *is defined as the continuum of relevant biogeographic and socioeconomic outcomes (e.g., geographical or technological capability factors)* $s \in [0,z]$ *that might be beneficial for innovative outcomes. The parameter* $0 \le \beta \le 1$ *measures the impact of biogeographic/socioeconomic factors on the probability of successful innovations, and* $(1-\beta)$ *is the corresponding impact for the prevalence of novelty-seeking traits. Assuming an equal distribution of the type of behavioral outcomes and the continuum of socioeconomic factors, that is,* $q(u) = 1/\nu$ *for all* ν, *and* $h(s) = 1/z$ *for all* z, *the parametrization of* $p(z,\nu)$ *takes the following form:*

$$p(z,\nu) = z^{1-\beta}\nu^{\beta}. \qquad (26)$$

In a similar vein, I assume that the economy generates research projects according to the following constant returns-to-scale function for the research production technology:

ASSUMPTION 4. *Without loss of generality, the research production technology* $\phi: \mathbb{R}_+^2 \to \mathbb{R}_+$ *can be parameterized according to the following expression:*

$$\phi(\ell_t^R, R) = \ell_t^R R^{\varphi}, \qquad (27)$$

where $R \ge 0$ *indicates a research-specific productivity parameter, and* $0 \le \varphi \le 1$ *is a parameter measuring the impact of R&D resources on research activity.*[3]

A complete graphical illustration of the model's equilibrium dynamics is provided in Figure 2. For illustrative purposes, I use the values shown in Table 2 for

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

TABLE 2. Summary of the model parameterization

Parameter	Notation	Value
ψ	Potential market distortion in the research sector	0.75
R	R&D resources	0.25
z	Biogeographical and socioeconomic factors	0.25
ν	Prevalence of novelty-seeking traits	0.25
γ	Productivity factor	1.50
α	Output elasticity *w.r.t.* intermediate goods	0.50
β	Probability of success parameter	0.50
φ	Research technology parameter	1.00
L	Total labor/population size	15.00
r	Real interest rate	0.10

model parametrization. It is worth mentioning that the inferences drawn from the model's equilibrium dynamics remain qualitatively unaffected by this particular choice of the model's parameterization.

Figure 2a depicts the functions $c(\ell_t^R)$ and $b(\ell_{t+1}^R)$. Now it becomes apparent that a higher amount of future research will discourage current research employment because it will result in a decrease of the marginal benefit and an increase in the marginal cost of research. Figure 2b depicts the forward-looking difference equation $\ell_t^R = \Psi(\ell_{t+1}^R)$. To illustrate the equilibrium path of the model, consider a situation where the economy initially starts at point $\overline{\ell_0^R}$. The corresponding value of future employment that satisfies the equilibrium condition in (24) is $\overline{\ell_1^R}$. The value of ℓ^R that would be chosen in the next period is given by $\overline{\ell_2^R}$. A closer look at Figure 2 shows that the economy approaches the static equilibrium $\widehat{\ell_{DE}^R}$ in a *DE* in a counterclockwise spiral setting starting at $\overline{\ell_0^R}$. The pair $(0, \overline{\ell_{ng}^R})$ constitutes a two-cycle "no-growth trap" equilibrium. In such a situation, the prospect of high future research completely discourages current research. The following central proposition describing the equilibrium dynamics of the model holds:

PROPOSITION 2. *There exists a sequence $\{\ell_t^R\}_{t=0}^{\infty}$ satisfying condition (25) for all $t \geq 0$, which constitutes a perfect foresight equilibrium (PFE). A static equilibrium corresponds to a PFE satisfying the condition $\ell_t^R = \ell_{t+1}^R = \ell_{t+2}^R = \ldots$ for all $t \geq 0$. The stationary equilibrium is graphically illustrated by the point at which the $\Psi(\ell_{t+1}^R)$ curve intersects the 45-degree line, as shown in Figure 2b.*

In the rest of this section, I provide a discussion of the stationary equilibrium of the model and derive some central comparative static results regarding the model's key parameters.

Substituting equations (26) and (27) in equation (24) and defining the condition of a stationary equilibrium in which the flow of research employment is constant across innovation intervals, that is, $\ell_t^R = \ell_{t+1}^R \equiv \widehat{\ell_{DE}^R}$, yields the following steady-state solution of research employment in a DE:

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

$$\widehat{\ell}_{DE}^{R} = \frac{(1-\psi)R^{\varphi}z^{1-\beta}\nu^{\beta}\gamma\left(\frac{1-\alpha}{\alpha}\right)L - r}{(1-\psi)R^{\varphi}z^{1-\beta}\nu^{\beta}\left[1+\gamma\left(\frac{1-\alpha}{\alpha}\right)\right]}. \tag{28}$$

Given the solution for the steady-state flow of research employment, the Poisson arrival rate of successful innovations in a DE is given by

$$\widehat{\lambda}_{DE}\left(\psi, \widehat{\ell}_{DE}^{R}, R, z, \nu\right) = \frac{(1-\psi)R^{\varphi}z^{1-\beta}\nu^{\beta}\gamma\left(\frac{1-\alpha}{\alpha}\right)L - r}{\left[1+\gamma\left(\frac{1-\alpha}{\alpha}\right)\right]}. \tag{29}$$

This expression is strictly positive if the condition $(1-\psi)R^{\varphi}z^{1-\beta}\nu^{\beta}\gamma(\frac{1-\alpha}{\alpha})L/r > 1$ holds. Standard comparative static analysis yields the following proposition:

PROPOSITION 3. *The steady-state Poisson arrival rate in a DE, $\widehat{\lambda}_{DE}(\psi, \widehat{\ell}_{DE}^{R}, R, z, \nu)$, positively depends on (a) the amount of R&D resources R, (b) the biogeographic/socioeconomic environment z, (c) the prevalence of novelty-seeking traits ν, (d) the productivity parameter γ, and (e) the total labor/population size L. It negatively depends on (a) potential market distortions in the research sector ψ and (b) the real interest rate r.*

2.5. Balanced Growth Path

For completeness, this section provides a derivation of the balanced growth path in a DE. Real output in the static equilibrium during innovation interval t is given by

$$y_t = A_t\left(L - \widehat{\ell}_{DE}^{R}\right)^{\alpha}. \tag{30}$$

Using the condition $A_{t+1} = \gamma A_t$, it follows that real output during innovation interval $(t+1)$ equals real output during interval t up to a productivity factor γ, that is,

$$y_{t+1} = A_{t+1}\left(L - \widehat{\ell}_{DE}^{R}\right)^{\alpha} \equiv \gamma y_t. \tag{31}$$

The stochastic process that drives real output is highly non-stationary. In order to derive the average growth rate during a unit time interval, knowledge about the expected number of innovations between two discrete points in time is needed. Note that the time path of real output can be written as

$$y(\tau+1) = \gamma^{t_{\Delta(\tau+1,\tau)}} y(\tau) \quad \textit{for}\ \tau = \{0, 1, 2, \ldots\}, \tag{32}$$

where $t_{\Delta(\tau+1,\tau)}$ is the number of innovations during the unit time interval $\Delta(\tau+1,\tau)$, that is, between period $(\tau+1)$ and τ. The log-linearization of equation (32) yields

$$\ln y(\tau+1) = \ln y(\tau) + \varepsilon\left[\Delta\left(\tau+1,\tau\right)\right] \quad \textit{for}\ \tau = \{0, 1, 2, \ldots\}, \tag{33}$$

where $\varepsilon[\Delta(\tau+1,\tau)] = t_{\Delta(\tau+1,\tau)} \ln\gamma$. From the theoretical framework, it follows that the number of innovations during the unit time interval follows a Poisson distribution with arrival rate $\widehat{\lambda}_{DE}(\psi, \widehat{\ell}_{DE}^{R}, R, z, \nu)$, that is,

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

$$t_{\Delta(\tau+1,\tau)} \equiv \frac{\varepsilon\,[\Delta\,(\tau+1,\tau)]}{\ln\gamma} \sim Poisson\big[\widehat{\lambda}_{DE}\big(\psi,\widehat{\ell}^{R}_{DE},R,z,\nu\big)\big] \text{ } for\ \tau=\{0,1,2,\ldots\}, \quad \textbf{(34)}$$

where $\mathbb{E}[t_{\Delta(\tau+1,\tau)}] = \mathbb{V}[t_{\Delta(\tau+1,\tau)}] = \widehat{\lambda}_{DE}(\psi,\widehat{\ell}^{R}_{DE},R,z,\nu)$. Combining equations (33) and (34), the economy's expected growth rate (*EGR*) and its corresponding variance (*VGR*) for $\tau=\{0,1,2,\ldots\}$ equals

$$EGR \equiv \mathbb{E}[\ln y(\tau+1) - \ln y(\tau)] = \mathbb{E}[\varepsilon[\Delta(\tau+1,\tau)]] = \widehat{\lambda}_{DE}\left(\psi,\widehat{\ell}^{R}_{DE},R,z,\nu\right)\ln\gamma, \quad \textbf{(35)}$$

$$\begin{aligned} VGR &\equiv \mathbb{V}[\ln y(\tau+1) - \ln y(\tau)] = \mathbb{V}[\varepsilon[\Delta(\tau+1,\tau)]] \\ &= \widehat{\lambda}_{DE}\left(\psi,\widehat{\ell}^{R}_{DE},R,z,\nu\right)(\ln\gamma)^2. \end{aligned} \quad \textbf{(36)}$$

The following proposition summarizes the impact of the various parameters on the economy's expected growth rate:

PROPOSITION 4. *The economy's steady-state growth rate in a DE, $\widehat{\lambda}_{DE}(\psi,\widehat{\ell}^{R}_{DE},R,z,\nu)\ln\gamma$, positively depends on (a) the amount of R&D resources R, (b) the biogeographic/socioeconomic environment z, (c) the prevalence of novelty-seeking traits ν, (d) the productivity parameter γ, and (e) the total labor/population size L. It negatively depends on (a) potential market distortions in the research sector ψ, and (b) the real interest rate r.*

2.6. Policy Experiment: A Rise in the Prevalence of Novelty-Seeking Traits

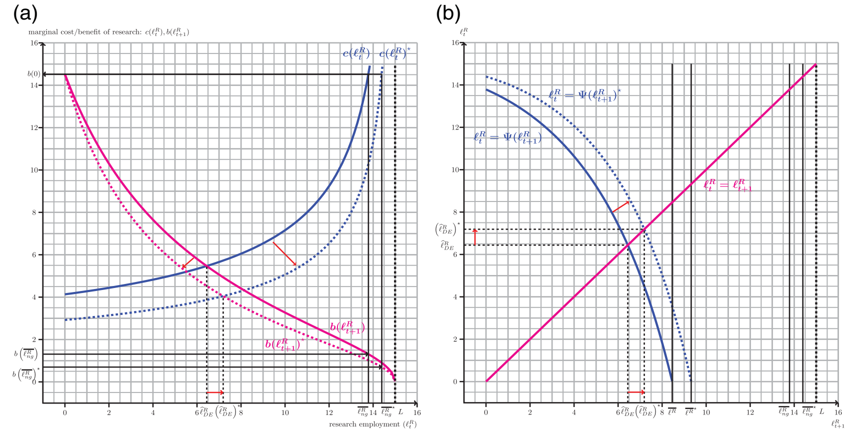
In this section, I examine the effects of a rise in the prevalence of novelty-seeking traits ν (e.g., due to a controlled migration policy) on the number of innovations and, thus, economic growth in the decentralized equilibrium. The model's central prediction is summarized in the following proposition:

PROPOSITION 5. *The economy's steady-state Poisson arrival rate (expected number of innovations) positively depends on the prevalence of novelty-seeking traits ν. Standard comparative static analysis reveals the following central result with respect to the parameter ν*

$$\frac{\partial\widehat{\lambda}_{DE}\left(\psi,\widehat{\ell}^{R}_{DE},R,z,\nu\right)}{\partial\nu} = \frac{(1-\psi)R^{\varphi}\beta(z/\nu)^{1-\beta}\gamma\left(\frac{1-\alpha}{\alpha}\right)L}{\left[1+\gamma\left(\frac{1-\alpha}{\alpha}\right)\right]} > 0. \quad \textbf{(37)}$$

The role of novelty-seeking traits ν in achieving $\partial\widehat{\lambda}_{DE}(\psi,\widehat{\ell}^{R}_{DE},R,z,\nu)/\partial\nu > 0$ is not obvious at first glance. On the one hand, a rise in the parameter ν encourages current research due to a decrease in the marginal cost of research $c(\ell^{R}_{t})$. As a result, the Poisson arrival rate and thus the number of innovations will increase as each unit of research becomes more productive. On the other hand, a higher prevalence of novelty-seeking traits ν will result in a decrease of the marginal benefit of research $b(\ell^{R}_{t+1})$, resulting in a higher rate of "creative destruction" in future periods. This circumstance will discourage firms from investing more in current

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press



The relationship between future and current research

Depicting the functional relationship $\ell_t^R = \Psi\left(\ell_{t+1}^R\right)$

FIGURE 3. Policy experiment: A rise in the prevalence of novelty-seeking traits ($\triangle \nu > 0$).

research because of the higher probability of being replaced by the next innovating firm that would result in a complete loss of the monopolist profit flow. Since the first effect outweighs the second, the overall impact on the expected number of innovations remains positive. Figure 3 nicely illustrates the corresponding shifts of the model's key functional representations and the derivation of the new steady-state research employment $(\widehat{\ell_{DE}^R})^\star$ resulting from a rise in the parameter ν. From the above discussion about the economy's balanced growth path, it is apparent that any rise in the model's Poisson arrival rate directly translates into higher economic growth mediated by a larger number of innovations. We thus obtain the following proposition regarding the long-run effect of an increase in the parameter ν on the economy's expected growth rate:

PROPOSITION 6. *The economy's steady-state economic growth rate positively depends on the prevalence of novelty-seeking traits ν. Standard comparative static analysis reveals the following central result with respect to the parameter ν:*

$$\frac{\partial EGR}{\partial \nu} = \frac{\partial \widehat{\lambda}_{DE}\left(\psi, \widehat{\ell_{DE}^R}, R, z, \nu\right)}{\partial \nu} \ln \gamma = \frac{(1-\psi)R^{\varphi}\beta(z/\nu)^{1-\beta}\gamma\left(\frac{1-\alpha}{\alpha}\right)L}{\left[1+\gamma\left(\frac{1-\alpha}{\alpha}\right)\right]} \ln \gamma > 0. \tag{38}$$

3. ECONOMETRIC SPECIFICATION AND ESTIMATION METHODOLOGY

The theoretical framework predicts that the prevalence of novelty-seeking traits in society is conducive to innovation activity in the aggregate economy, which is consistent with the hypothesis that such traits facilitate the creation of knowledge through beneficial economic attitudes related to risk-taking, creativity, and explorative behavior. Given equation (29), I derive a reduced-form regression equation of the Schumpeterian growth model, describing the number of innovations at decentralized equilibrium. The many parameters involved in the theoretical model can thereby be mapped to a regression equation that can be estimated using appropriate econometric estimators. The arrival of innovations per unit time interval as the main dependent variable typically takes on non-negative values with no theoretical upper bound. In the following, I show that the appropriate econometric estimator to model the arrival of innovations per unit time is the basic Poisson regression model where the dependent variable follows a Poisson distribution.

From the stochastic process in (14) and equation (29), it follows for the conditional mean function regarding the expected number of innovations T of a particular country i:

$$\mathbb{E}\left[T=t_i|\psi_i, \widehat{\ell}^R_{iDE}, R_i, z_i, \nu_i\right] = \widehat{\lambda}_{iDE}\left(\psi_i, \widehat{\ell}^R_{iDE}, R_i, z_i, \nu_i\right), \quad \textbf{(39)}$$

where $\widehat{\ell}^R_{iDE}$ and $\widehat{\lambda}_{iDE}$ refer to the steady-state solution of research employment and the Poisson arrival rate of successful innovations of country i in a DE, respectively. Even though equation (39) is a highly non-linear function, the model's parameters have a meaningful interpretation. To see this, first take the natural logarithm of both sides of the equation:

$$\ln \mathbb{E}\left[T=t_i|\psi_i, \widehat{\ell}^R_{iDE}, R_i, z_i, \nu_i\right] = \ln\left[\frac{(1-\psi_i)R_i^{\varphi} z_i^{(1-\beta)} \nu_i^{\beta} \gamma_i \left(\frac{1-\alpha}{\alpha}\right) L_i - r_i}{\left[1+\gamma_i\left(\frac{1-\alpha}{\alpha}\right)\right]}\right]. \quad \textbf{(40)}$$

Under the assumption that r, the real interest rate, is sufficiently small, it can be shown that the logarithm of the expected value is linear in the model's parameters:

$$\begin{aligned}\ln \mathbb{E}\left[T=t_i|\psi_i, \widehat{\ell}^R_{iDE}, R_i, z_i, \nu_i\right] \approx{}& \ln L_i + \ln(1-\psi_i) + \varphi \ln R_i + (1-\beta)\ln z_i \\ &+ \ln\left(\frac{1-\alpha}{\alpha}\right) + \ln\left[\frac{\gamma_i}{1+\gamma_i\left(\frac{1-\alpha}{\alpha}\right)}\right] + \beta \ln \nu_i,\end{aligned} \quad \textbf{(41)}$$

as $r \to 0$. In principle, this log-linearized model can be estimated by ordinary least squares. In this case, the regression coefficient β would correspond to an elasticity with respect to the prevalence of novelty-seeking traits ($\nu > 0$). However, given the data generating process for T, linear regression models might be unable to fully take into account the non-negative counts of innovations per year: $\{0, 1, 2, \ldots\}$. It is obvious that a linear regression model ceases to work in cases where T can take values of zero until an adjustment of ln(0) is made. Furthermore, linear regression models cannot ensure that the predicted value of T is positive for any combination of the explanatory variables and the estimated regression parameters. Thus, for discrete count data, it seems appropriate to model $\mathbb{E}[T=t_i|\psi_i, \widehat{\ell}^R_{iDE}, R_i, z_i, \nu_i]$ directly rather than its log-linear representation using a functional form that ensures positive values of T over the entire range of the model's explanatory variables and parameter values. In this case, the population regression model of $\mathbb{E}[T = t_i|\psi_i, \widehat{\ell}^R_{iDE}, R_i, z_i, \nu_i]$ is given by the exponential function of equation (41)

$$\mathbb{E}\left[T = t_i|\psi_i, \widehat{\ell}^R_{iDE}, R_i, z_i, \nu_i\right] = L_i \exp\left[\boldsymbol{x}_i'\boldsymbol{\theta} + \beta \ln \nu_i\right], \qquad \textbf{(42)}$$

where $\boldsymbol{x}_i'\boldsymbol{\theta} = \ln(1-\psi_i) + \varphi \ln R_i + (1-\beta)\ln z_i + \ln(\frac{1-\alpha}{\alpha}) + \ln[\frac{\gamma_i}{1+\gamma_i(\frac{1-\alpha}{\alpha})}]$.[4] This non-linear model is estimated by the *maximum likelihood* method under the assumption that T is drawn from a Poisson distribution. It is straightforward to see that equation (42) corresponds to a basic Poisson regression model in which the conditional mean function shown in (39) is parameterized with the exponential function as $r \to 0$.[5] Note that L_i is treated as the *exposure* variable, a terminology frequently used in Poisson regression models. This accounts for a type of heteroscedasticity in the model related to the population size of the country. Specifically, it seems natural to assume that the total number of innovations per year in a particular country might be related to the country's population size. Constraining the parameter estimate of $\ln L_i$ to equal 1 accounts for this type of heterogeneity across countries. In this regard, the parameter estimate β reflects the percentage change in the number of innovations *per capita* of a one percentage change in the prevalence of novelty-seeking traits in society ν.

Because the main objective is to consistently estimate the effect of the prevalence of novelty-seeking traits on innovation activity, the Poisson regression model accounts for a full set of additional country-specific factors, as summarized by the term $\boldsymbol{x}_i'\boldsymbol{\theta}$. The parameterization of these factors in the empirical analysis is accomplished with a full set of biogeographic, historical, and socioeconomic control variables. The line of reasoning for the inclusion of these factors in the regression model is discussed in more detail further below in the data description section. For example, the parameter z in the model is approximated through the inclusion of a full set of microgeographic (e.g., distance to major markets), land productivity (e.g., percentage arable land area), climate (e.g., mean temperature and precipitation), and health environment factors (e.g., percentage of the population living in tropics). The inclusion of biogeographic controls rules out the possibility that the relationship between knowledge creation and the country-level measure of novelty-seeking traits simply reflects the issue of geographic proximity. Moreover, differences in biogeographic conditions might themselves explain variations in knowledge creation across countries. For example, the factors responsible for low labor productivity in areas located in the tropics (e.g., health outcomes, life expectancy, and the formation of human capital in malaria-prone regions) appear to play a direct and important role in the creation of knowledge in society. Additionally, geographic factors affect knowledge creation indirectly, through past institutional, cultural, and economic events and processes whose effects persist to the present day.

In addition to this standard set of country-level controls, I further assess the sensitivity of the main results to the inclusion of cultural (e.g., religious background), historical (e.g., colonial heritage), diversity (e.g., genetic), technological frontier (e.g., linguistic distance to the United States), technological and human capability (e.g., R&D expenditures as percentage of GDP), as well as economic preference (e.g., risk-taking behavior) factors. In summary, the choice of relevant variables is intended to rule out two main considerations in the empirical analysis: first, the possible existence of confounding factors and, second, the possible omission of key determinants of knowledge creation. These are discussed in greater detail in the next section.

4. DATA AND VARIABLES

In this section, I provide a detailed discussion of the set of variables employed in the empirical analysis. First of all, I introduce the main dependent variable that indicates the extent of knowledge creation or innovation activity in the research sector. I then proceed to provide a detailed discussion regarding the human DRD4 exon III gene that population geneticists recurrently linked to the human phenotype of novelty-seeking behavior in individuals. Furthermore, I discuss possible strengths and limitations of the use of genetic biomarkers for the assessment of complex personality traits in general and novelty-seeking behavior in particular. Finally, this section further presents a variety of biogeographic, climatic, socioeconomic, historical, and cultural variables that have attracted considerable attention in the development economics literature that appear to be equally important for both the extent of knowledge creation in society and the distribution of novelty-seeking traits across countries.[6]

4.1. Main Dependent Variable: Number of Scientific and Technical Journal Articles

An empirical investigation of the determinants of knowledge creation requires a proper definition of research and development (R&D) activities allowing them to be identified and measured in the aggregate economy. The OECD's Frascati Manual provides a guideline for the measurement of scientific, technological, and innovation activities. According to this guideline, R&D activities are defined as "[...] creative and systematic work undertaken in order to increase the stock of knowledge—including knowledge of humankind, culture and society—and to devise new applications of available knowledge" [(OECD, 2015, pp. 44–45)]. This definition covers three types of R&D activities—basic research, applied research, and experimental development—depending on their applicability of achieving specific goals (e.g., creation of basic knowledge, products, or processes). The output of the research activities is usually published in scientific or technical journal articles. Therefore, the main dependent variable employed in this study and throughout the empirical analysis refers to the number of scientific and technical journal articles that have been published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. I use this variable to measure the level of knowledge creation in society.[7]

4.2. Main Explanatory Variable: DRD4 Exon III 2- and 7-Repeat Allele Frequency

A main challenge in the empirical analysis pertains to the difficulty of modeling and measuring a latent human phenotype such as novelty-seeking behavior that encompasses a wide variety of possible behavioral dimensions. According to the three-dimensional personality questionnaire developed by Cloninger (1987), novelty-seeking individuals are characterized as impulsive, exploratory, fickle, excitable, quick-tempered, and extravagant. Scientific evidence from molecular genetics research is a promising strategy to overcome data limitation issues. Specifically, Faraone et al. (2014) suggest the use of a set of genetic biomarkers that are useful to predict the development of personality and/or behavioral traits. In contrast to self-reported personality outcomes, genetic biomarkers offer a promising strategy to improve the assessment of complex human phenotypes, as the latter is unlikely to be confounded by individual level characteristics (e.g., age, gender, and socioeconomic status). The narrative of this approach is that the prevalence of specific genetic biomarkers helps to predict the tendency of people to display certain personality outcomes.

Novelty-seeking-related behavioral outcomes (e.g., exploratory, excitability, and impulsivity) are natural candidates to be biologically influenced by genetic variants that are located in brain regions associated with the human dopamine system [Paterson et al. (1999)]. One such candidate gene is the dopamine D4 receptor gene (DRD4), which consists mainly of four encoded regions, called exons by population geneticists, of which the third exon shows the most extensive polymorphism in allelic variants.[8] This polymorphism occurs as a 48-base pair (48-bp) variable number of tandem repeats, ranging from 2- to 11-repeats, with the 2-repeat, 4-repeat, and 7-repeat being the primary allelic variants found in the human genome across populations [Chen et al. (1999); Gören (2016)].

Dopamine functions as an important neurotransmitter that regulates intracellular transmission of information between synaptic clefts in the brain. Dopamine binds to the post-synaptic receptors and triggers a cell response that eventually results in a physiological change (e.g., control of locomotion, reward, and emotional stability) [Civelli et al. (1991)]. Thus, any dysregulation associated with dopamine release and/or receptor binding between synaptic clefts in the brain would be associated with behavioral malfunctioning.

A striking observation regarding the DRD4 exon III gene is that the various primary allelic variants (e.g., the 2-repeat, 4-repeat, and 7-repeat) exhibit significant physiological differences to dopamine releases [Van Tol et al. (1991, 1992); Lichter et al. (1993)]. Specifically, it has been reported that the 2-repeat and

7-repeat allele variants show a suboptimal blunted response to dopamine release relative to the ancestral 4-repeat allele variant [Asghari et al. (1995); Wang et al. (2004)]. It has been hypothesized that this kind of blunted response to dopamine stimulation requires elevated dopamine levels between synaptic clefts in the brain that makes some people more sensitive to environmental stimuli, finally affecting the development of particular phenotypic characteristics such as novelty-seeking behavior [Swanson et al. (2000)].

Consistent with the dopamine hypothesis of novelty-seeking behavior, two influential candidate gene association studies identified a significant relationship between DRD4 exon III polymorphism (in particular the presence of the 7-repeat allele variant) and self-reported novelty-seeking test score ratings in two independent population samples [Benjamin et al. (1996); Ebstein et al. (1996)].

An indirect approach to examine the predictive power of DRD4 exon III polymorphism on the development of novelty-seeking behavior would be to focus on extreme personality outcomes such as ADHD, a neuropsychological disorder that shares many behavioral dimensions (e.g., exploratory and impulsive behavior) usually found among individuals with high novelty-seeking test score ratings [Ebstein (1997)]. In accordance with the theory of dopamine dysregulation of ADHD proposed by LaHoste et al. (1996), a series of meta-analysis report a strong and robust positive association between DRD4 exon III polymorphism and ADHD-related behavioral symptoms inattention and/or hyperactivity-impulsivity [Faraone et al. (2005); Bobb et al. (2006); Gizer et al. (2009); Wu et al. (2012)]. This finding provides evidence that novelty-seeking behavior constitutes a behavioral symptom that is present with different degrees of intensity among individuals.

Nevertheless, it is worth mentioning to indicate possible weaknesses of candidate gene association studies. In particular, the genetic architecture of a complex personality trait such as novelty-seeking is most likely to be polygenic in nature. Based on evidence from large-scale genome-wide association studies, it is conceivable that novelty-seeking behavior is particulary determined by many genes or combination of genes, each contributing only a small fraction to the total variation of observed novelty-seeking test score outcomes [Chabris et al. (2012)].

Furthermore, the pleiotropic effects of DRD4 exon III polymorphism have been examined in additional candidate gene association studies of other personality traits than novelty-seeking behavior. Given this uncertainty that underlies the interpretation of the DRD4 exon III polymorphism, one should be cautious in interpreting the main findings as providing full support for novelty-seeking behavior, and instead search for alternative channels that might confound the relationship between knowledge creation and the prevalence of novelty-seeking traits in society. For example, several studies reported a possible association between DRD4 exon III polymorphism and risk-taking, patience, altruism *vis-à-vis* selfishness, creativity, fairness, and other pro-social behaviors such as trust [Kuhnen and Chiao (2009); Dreber et al. (2009); Carpenter et al. (2011); Jiang et al. (2013); Mayseless et al. (2013); Cochran and Harpending (2009)]. Thus, it seems that besides novelty-seeking behavior, DRD4 exon III polymorphism is capturing other but related personality traits that are of economic interest.[9]

The study in Gören (2017) provides DRD4 exon III 2- and 7-repeat allele frequencies in a large sample of 181 countries across the world, which I use as the key explanatory variable in the empirical analysis on the prevalence of novelty-seeking and/or related personality traits in society. These country-level genetic measures were constructed by matching the entire distribution of ethnic groups in the Alesina et al. (2003) ethnicity data to the DRD4 exon III population genome data in Gören (2016) based on the phylogenetic relationship among ethno-linguistic groups, as indicated by the *Ethnologue* database [Global Mapping International (2010)]. Genetic information of the sampled populations were compiled from a large number of molecular genetic studies, reporting DRD4 exon III allele frequencies of healthy (non-psychiatric) individuals [Gören (2016)]. Unlike self-reported measures of individual economic attitudes, a limited number of people are sufficient to provide a representative picture of the overall genetic composition of entire populations. For example, the Human Genome Diversity Cell Line Panel from the Human Genome Diversity Project-Centre d'Etude du Polymorphisme Humain (HGDP-CEPH) employs genomic data on 1,050 individuals to infer the within-genetic composition of 52 populations across the world [Cann et al. (2002)]. The same rationale applies to the population genome data compiled by Gören (2016) to provide a reliable database of the worldwide distribution of population-specific DRD4 exon III allele frequencies across a large number of 120 populations located in Africa, the Americas, Asia, Europe, the Middle East, and Oceania.[10]

It is worth mentioning that the proposed country-level DRD4 exon III 2- and 7-repeat allele frequency measure should not be misinterpreted in a way that a higher frequency of these allelic variants biologically determines the development of a complex personality trait such as novelty-seeking behavior. Instead, the argument is that those societies might have a higher probability that a larger continuum of novelty-seeking-related behavioral outcomes might be observed in the relevant settings [Plomin et al. (2013)]. In this regard, the country-level DRD4 exon III polymorphism should be interpreted as a simple proxy variable that is correlated with the latent personality trait of novelty-seeking behavior in society.

4.3. Additional Control Variables: Biogeographical, Historical, and Socioeconomic Factors

Biogeographic factors. An empirical analysis of the determinants of scientific knowledge creation should include a full set of microgeographic, land productivity, climatic, health, and regional factors. For example, geographic proximity to technologically advanced countries might facilitate knowledge diffusion across national borders, which in turn could positively affect the country's own rate of technological progress. Furthermore, it has been reported that the country's

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

natural endowment with mineral resources might result in widespread corruption and rent-seeking activities that ultimately discourage the accumulation of individual human capital in society [Sachs and Warner (2001)]. The heavy disease burden associated with tropical areas (e.g., prevalence of malaria falciparum) is another important determinant of cross-country differences in standards of living due to its potential detrimental impact on labor productivity [Gallup et al. (1999)].

Besides providing a source of cross-country variation in scientific knowledge creation, biogeographic controls fulfill another important purpose in the empirical analysis. Recent evidence suggests that local biogeographic conditions contribute significantly to between-population variation of DRD4 exon III allele variants [Gören (2016)]. Therefore, the inclusion of a large set of biogeographical controls in the regression model effectively rules out endogeneity concerns that the proposed association between knowledge creation and the prevalence of novelty-seeking traits solely reflects the issue of geographic proximity.

Diversity factors. The issue of ethnic, linguistic, religious, and even genetic diversity is the subject of intense scientific debate on the potential costs and benefits of a diverse society on socioeconomic outcomes such as redistribution, provision of public goods, economic growth, and the incidence of conflicts [Alesina et al. (2003); Desmet et al. (2012); Gören (2014)]. On the one hand, a more diverse society may exhibit a wide range of human skills that might be complementary in an increasingly complex production environment. In theory, teams composed of diverse members may outperform homogenous ones due to increased problem-solving capabilities of the former group [Hong and Page (2001, 2004)]. In contrast, the potential costs of a diverse society on aggregate productivity are associated with higher communication costs, lower interpersonal trust, and conflicts. These factors appear equally important for the level of innovation activity as it may hamper the diffusion of knowledge in society.

Recent evidence on the potential benefits and costs of genetic diversity on aggregate productivity suggests that this issue affects economic development quite differently than, for example, ethnic diversity [Ashraf and Galor (2013)]. Several mechanisms have been reported that might be consistent with the non-monotonic inverted-U relationship between genetic diversity and contemporary economic development. It has been suggested that the potential costs of genetic diversity are associated with lower interpersonal trust and higher conflict incidence in society [Ashraf and Galor (2013); Arbatli et al. (2020)]. More importantly, in regard to the research focus of this paper, the study in Ashraf and Galor (2013) shows that the potential benefits of genetic diversity work through increased innovation activity, as indicated by the number of scientific and technical journal articles per capita in a cross section of countries.

Historical factors. The enduring legacy of historical factors has attracted considerable attention in the field of development economics. Specifically, it has been reported that the percentage of Europeans in countries during the early stages of colonization contributed significantly to local economic development [Acemoglu et al. (2001, 2002)]. This finding is interpreted to suggest Europeans developed specific kinds of political institutions, human capital factors, technology, and culture that were particularly conducive to the process of economic development [Glaeser et al. (2004); Easterly and Levine (2016)].

Additional mechanisms by which a country's colonial history might affect scientific knowledge creation relate to the issue of knowledge diffusion between former colonized countries and the former colonizing powers. For example, countries that were once British colonies might find it easier to absorb knowledge and technology from Great Britain. Related to this aspect is the issue of whether the country's official language is English. This definition covers former colonies that adopted English as their official language in school, business, and public transactions. The inclusion of this variable effectively controls for two distinct but important aspects in the empirical analysis. First, this variable is intended to capture the country's linguistic proximity to the language of scientific communication (e.g., English) which may be particularly capable of absorbing global knowledge stocks, among other things. Second, bibliographic information from the Science Citation Index and Social Sciences Citation Index databases may be biased toward scientific and technical articles published in English-speaking journals. This may underestimate the extent of scientific knowledge creation in non-English-speaking countries.

Another focus of historical research is on the legal traditions characterizing a country's political system, which capture the extent of the government's interventions into political and social life [La Porta et al. (1999)]. The protection of individual rights is certainly an important prerequisite for the freedom of expression and individual self-realization, which substantially affect the creation and diffusion of knowledge in society.

To consider cultural factors as another aspect of history, the baseline specification further includes a full set of a country's major religions (i.e. percentage of protestants, catholics, and muslims). It has been argued for quite some time that the kind of cultural norms and beliefs frequently associated with adherents to Protestantism (e.g., higher literacy rate) resulted in the so-called "Protestant work ethic", which was particularly beneficial for the process of economic development [Weber (1958)]. The promotion of literacy among Protestants and their relative emancipation from the church may have contributed significantly to the rise of secular knowledge in society.

Technological frontier factors. In research on the human barriers to knowledge diffusion, Spolaore and Wacziarg (2009) have shown that measures of relative genetic distance to the technological frontier (e.g., the United States) appear to have a statistically significant and economically sizeable effect on cross-country differences in income per capita. The authors interpret their results as evidence that genetic distance between populations captures differences in culture, norms, and beliefs that affect the diffusion of technology across societies.

In other words, countries that are genetically similar to the frontier society may find it easier to adopt recent technological advances that in turn positively affect their own economic development [Spolaore and Wacziarg (2012)]. This finding is consistent with micro-level evidence on the determinants of international patent citations, highlighting the importance of knowledge diffusion across countries through ethnic scientific networks [Kerr (2008)].

In the empirical analysis, I consider various measures of genetic, linguistic, and religious distance from the technological frontier (i.e. United States, OECD, and Neo-Europe). Even though the correlation between the various technological distance measures is quite high, its implications for cross-country knowledge diffusion through long-term genealogical, linguistic, and religious differences may differ. For example, cultural differences in habits, norms, and beliefs may result in lower trust and communication between populations that in turn prevent the diffusion of knowledge and ideas across countries, even though the two populations are genetically indistinguishable from each other (e.g., German and French).

Technological and human capability factors. The creation and adoption of new scientific knowledge require specific investments in a country's technological and human capability resource base. This definition is similar to the notion of absorptive capacity initially proposed by Cohen and Levinthal (1990). In theory, models of endogenous economic growth already suggest that cross-country differences in the rate of technological progress are attributable to differences in the country's human capital and R&D resource base [Romer (1990)]. Even though the identification and measurement of the relevant technological and human capability factors is a difficult task to accomplish in empirical work, the survey presented by Fagerberg et al. (2010) provides a useful guideline for identification of the relevant factors. Specifically, a country's level of economic development might be used as an overall measure for standards of living that are highly correlated with other proximate factors of economic growth, such as investments in human capital, physical capital, and infrastructure quality, among others. Besides the use of GDP per capita in the empirical analysis, I employ additional variables to identify the specific channels that facilitate scientific knowledge creation in society (i.e. various proxy variables related to national R&D efforts, infrastructure quality, and the level of human capital skills).

Economic preference factors. An extensive body of research has investigated the importance of a set of economic preference measures and various outcomes at both the individual and country level. At the individual level, several studies in behavioral and experimental economics have emphasized the predictive power of time preferences for individual saving decisions, educational attainment, cognitive ability, and wealth inequality [Epper et al. (2020); Sutter et al. (2013); Dohmen et al. (2010)]. Another related strand of literature has posited that an individual's willingness to take risks is significantly related to more risk-taking economic behaviors, including self-employment, entrepreneurial activities, holding stocks, as well as risky health behaviors [Bonin et al. (2007); Dohmen et al. (2011); Viscusi and Hersch (2001); Falk et al. (2018)]. Finally, at the country level, empirical evidence suggests a positive correlation between social preferences (e.g., trust, altruism, and positive reciprocity) and country-level outcomes in the areas of development, entrepreneurship, voluntary donations, and civil conflicts [Knack and Keefer (1997); Algan and Cahuc (2013); Falk et al. (2018)]. Motivated by this line of research, I employ a set of economic preference factors as potential drivers of knowledge creation across countries. For this purpose, I use a recent database on the global variation in risk-taking, patience, positive and negative reciprocity, altruism, and trust [Falk et al. (2018)]. In this regard, I am able to assess the role of other behaviorial dimensions that might be correlated with the prevalence of novelty-seeking traits in order to rule out alternative explanations regarding the key hypotheses proposed in this study.

5. EMPIRICAL RESULTS

This section presents the first results on the relationship between the number of scientific and technical journal articles per 1000 people and the country-level DRD4 exon III 2- and 7-repeat allele frequency measure. I first present coefficient estimates of the Poisson regression model that include various biogeographic controls. Afterward, I test the sensitivity of the main findings to the issue of genetic, ethnic, linguistic, and religious diversity. This step of the empirical analysis is of particular importance in order to rule out concerns that the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure simply captures unobserved effects related to various aspects of diversity in society. Once the robustness of the main findings to the inclusion of the various diversity controls has been established, I proceed in testing the robustness of the baseline specification to the inclusion of additional historical, technological frontier, technological and human capability, and economic preference factors.

Including biogeographic factors. Table 3 presents coefficient estimates for the set of biogeographic controls. The results shown in column (1) assess the relationship between the number of scientific and technical journal articles and the natural logarithm of the country-level $DRD4^{R2R7}$ measure without including any country-level controls.

Even though the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure is positive, it is not statistically significant at conventional significance levels. Consistent with the theoretical framework, this result suggests that the influence of the country-level $DRD4^{R2R7}$ measure on scientific knowledge creation in society might be conditional on additional country-specific controls.

Hence, the results presented in column (2) include a set of microgeographic factors in the Poisson regression model. Once conditioning on this set of microgeographic controls, the estimated coefficient associated with the country-level

TABLE 3. DRD4 exon III 2- and 7-repeat allele frequency and number of scientific and technical publications (Biogeographic factors)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: Number of scientific and technical publications, t_i						
ln *DRD4*^{R2R7}	0.5392	1.0165***	1.0772***	1.2129***	1.2179***	1.5692***	1.6541***
(Ethnicity-Weighted)	(0.8090)	(0.2838)	(0.2794)	(0.2119)	(0.2319)	(0.2774)	(0.4440)
ln Absolute latitude		1.2208***	1.2443***	1.2977***	1.3671**	0.7150**	0.7490**
		(0.2773)	(0.2940)	(0.2708)	(0.6564)	(0.3463)	(0.3543)
ln Distance to major markets		−0.5037***	−0.5098***	−0.4779***	−0.4788***	−0.2724***	−0.1920*
		(0.0594)	(0.0602)	(0.0711)	(0.0712)	(0.0689)	(0.1117)
ln Distance to coast or river		−0.1392***	−0.1219**	−0.2421***	−0.2334**	0.0968	−0.0060
		(0.0514)	(0.0595)	(0.0639)	(0.1003)	(0.1054)	(0.1467)
ln Hydrocarbons per person			−0.0206	−0.0127	−0.0130	−0.0335	−0.0300
			(0.0316)	(0.0253)	(0.0248)	(0.0275)	(0.0318)
Percentage arable land area				−0.0172*	−0.0178	0.0137*	0.0145*
				(0.0104)	(0.0124)	(0.0082)	(0.0077)
Agricultural suitability				−0.3784	−0.3728	−1.8714***	−2.0300***
				(0.5922)	(0.6092)	(0.6076)	(0.5496)
Terrain roughness				0.4487	0.4342	1.1821	0.9279
				(1.4145)	(1.5120)	(1.0314)	(1.0860)
Temperature					0.0047	0.0530**	0.0545*
					(0.0351)	(0.0264)	(0.0297)
Precipitation					0.0000	0.0078*	0.0055
					(0.0033)	(0.0041)	(0.0050)

TABLE 3. Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population in area with malaria						−2.1300***	−2.1316***
(Share of total population)						(0.4883)	(0.6288)
Population in tropics						−0.6808	−0.3932
(Share of total population)						(0.7617)	(0.7536)
Population in temperate zones						1.6836**	2.1052**
(Share of total population)						(0.6895)	(1.0085)
Log − Likelihood value	−990,357.63	−207,631.72	−206,572.28	−182,422.71	−182,327.03	−105,399.03	−101,095.03
$Pseudo - R^2$	0.01	0.79	0.79	0.82	0.82	0.89	0.90
$Corr[t_i, \widehat{t_i}]^2$	0.24	0.95	0.95	0.94	0.94	0.98	0.98
Number of countries	129	129	129	129	129	129	129
Region fixed effects	No	No	No	No	No	No	Yes

Notes: The dependent variable is the average number of scientific and technical publications between the years 1981 and 2016. Estimation methodology: Poisson regression model estimated by the Poisson maximum likelihood estimator. The exposure variable is the country's total population size averaged between the years 1981 and 2016. The expression $Corr[t_i, \widehat{t_i}]^2$ refers to the squared correlation between the observed (t_i) and predicted ($\widehat{t_i}$) number of scientific and technical publications. See the main text for additional details. ln $DRD4^{R2R7}$ refers to the log of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure (ethnicity-weighted). ln *Absolute Latitude* is the log of the absolute value of a country's approximate centroid latitude in decimal degrees. ln *Distance to Major Markets* is the log of the minimum great-circle distance (in 1000 km) from the country's capital city to New York, Rotterdam, or Tokyo. ln *Distance to Coast or River* is the log of mean distance (in 1000 km) to the nearest ice-free coastline or sea-navigable river. ln *Hydrocarbons per Person* is the log of British thermal units per person of proven crude oil and natural gas reserves in 1993. *Percentage Arable Land Area* is the country's percentage arable land area. *Agricultural Suitability* is a geospatial indicator, ranging from 0 to 1, of land suitability for agriculture across 0.5 decimal degrees latitude × longitude grid cells in the area covered by each country. *Terrain Roughness* is a geospatial indicator that indicates the average absolute change in elevation values across contiguous 1 decimal degrees latitude × longitude grid cells in the area covered by each country. *Temperature* is the mean country's temperature (in degree celsius) during the period 1960 and 1990. *Precipitation* is the mean country's precipitation (in total millimeters per month) during the period 1960 and 1990. *Population in Area with Malaria* is the share of a country's population in 1995 residing in areas contracting with malaria falciparum. *Population in Tropics* is the share of a country's population in 1995 residing in tropics. *Population in Temperate Zones* is the share of a country's population in 1995 residing in temperate zones. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, America, Asia, and Europe. Constant term included but not shown. Robust standard errors are reported between parenthesis.

*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

$DRD4^{R2R7}$ variable increases substantially in magnitude and turns highly statistically significant at the 1% significance level. Note that the country's overall population size functions as the exposure variable in the Poisson regression model throughout all model specifications, so the estimated regression coefficients should be interpreted accordingly.[11] In particular, the point estimate suggests that increasing the country-level $DRD4^{R2R7}$ value of Poland (0.1992) to the level of the United States (0.2636) would, ceteris paribus, increase the number of scientific and technical journal articles *per capita* by about $1.0165 \times \ln(0.2636/0.1992) \times 100\% = 28.47\%$.[12] This change in the dependent variable is equivalent to moving Poland from the 77.52nd percentile to the 79.07th percentile in the distribution of the number of scientific and technical journal articles *per capita*.

The remaining coefficient estimates are all of the expected signs. For example, the positive coefficient associated with the country's approximate absolute latitude in decimal degrees is consistent with productivity-enhancing climate conditions usually found in higher latitude regions [Gallup et al. (1999)]. Moreover, geographic proximity to the world's three major technological regions (e.g., the United States, continental Europe, and Japan) appears of considerable importance for the creation of knowledge in society through, for example, knowledge diffusion.

In column (3), I include the logarithm of the country's proven mineral resources (i.e. oil and natural gas reserves expressed in British thermal units per capita in the year 1993) in the regression equation. It is generally acknowledged that countries rich in mineral resources perform differentially in economic terms due, for example, to widespread corruption and rent-seeking activities that might hinder the accumulation of knowledge in society [Sachs and Warner (2001)]. As expected, the estimated coefficient associated with the log of hydrocarbons per capita is negative, but statistically insignificant at conventional significance levels.

The estimates presented in column (4) investigate the sensitivity of the main results to the inclusion of various land productivity factors. Neither of these variables alter the main findings substantially. The estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure increases both in magnitude and statistical significance, suggesting that the previous model specification was partially confounded by the omission of land productivity factors. It is worth mentioning that the coefficient associated with the percentage of arable land area enters with a negative sign in the regression equation. This result is consistent with the hypothesis that the extent of scientific knowledge creation is on average lower in countries that place a greater emphasis on agricultural practices.

In column (5), the relationship between scientific knowledge creation and the country-level $DRD4^{R2R7}$ measure is also robust to the inclusion of climatic factors. The positive coefficient associated with the country's mean precipitation and temperature value, respectively, is consistent with the notion that knowledge creation is higher in temperate climatic zones, even though the robustness of this finding is sensitive to the particular model specification.

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

Next, the results presented in column (6) examine the sensitivity of the previous findings to the inclusion of health-related factors in the regression model. Again, the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure remains highly statistically significant at the 1% significance level. Furthermore, the coefficient estimate regarding the share of a country's population in areas with malaria enters with the expected negative sign, consistent with the heavy disease burden and the resulting detrimental effects on the accumulation of knowledge in society [Gallup et al. (1999)].

Finally, column (7) shows that the main findings are robust to the inclusion of region fixed effects (i.e. indicator variables that take a value of one if the respective country is located in Sub-Sahara Africa, America, Asia, or Europe and zero otherwise) in the Poisson regression equation.

The issue of diversity. In the following, I assess the sensitivity of the previous findings to the issue of genetic, ethnic, linguistic, and religious diversity. Specifically, given the fact that the country-level DRD4 exon III 2- and 7-repeat allele frequency measure is positively correlated with migratory distance from East Africa [Gören (2017)], it is conceivable that the estimated regression coefficient simply captures the impact of genetic diversity on scientific knowledge creation [Ashraf and Galor (2013)]. Furthermore, I examine the sensitivity of the main results to the inclusion of ethnic, linguistic, and religious diversity to rule out concerns that the empirical findings might be prone to unobserved factors related to various definitions of diversity in society. The corresponding results are shown in Table 4. The estimates presented in column (1) replicates the baseline model specification in Table 3, column (7), and are shown for comparison purposes.

In column (2), I include the measure of genetic diversity, predicted by migratory distance from East Africa, in the regression equation. To account for the fact that most countries today have populations from various regions of the world, an ancestry-adjusted version of this measure is employed based on population flow data since 1500 AD [Putterman and Weil (2010)]. The estimated regression coefficient associated with predicted genetic diversity is positive and statistically significant at the 10% significance level, which is in line with previous findings reported by Ashraf and Galor (2013). More importantly, the main finding regarding the positive association between the country-level $DRD4^{R2R7}$ measure and the number of scientific and technical journal articles *per capita* remains robust to the simultaneous inclusion of predicted genetic diversity into the same regression model, as indicated in column (3).

The estimates reported in columns (4) and (5) examine the sensitivity of the main results to the inclusion of migratory distance from East Africa. This model specification is intended to rule out the possibility that the main findings might be confounded by unobserved factors to which migratory distance from East Africa might be related [Ashraf and Galor (2013)]. Reassuringly, the regression coefficient associated with the log of the country-level $DRD4^{R2R7}$ measure remains of the expected positive sign and statistically significant at the 1% significance level.

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

TABLE 4. DRD4 exon III 2- and 7-repeat allele frequency and number of scientific and technical publications (Diversity factors)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Dependent variable: Number of scientific and technical publications, t_i											
ln *DRD4*^{R2R7}	1.6370***		1.9211***		1.8807***		1.6774***		1.6469***		1.3854***	1.7728***
(Ethnicity-weighted)	(0.4527)		(0.4222)		(0.4193)		(0.4205)		(0.4186)		(0.3254)	(0.3544)
Predicted genetic diversity		9.6694*	17.6866***									73.5519*
(Ancestry-adjusted)		(5.1432)	(6.6760)									(43.4716)
Migratory distance from East Africa				−0.0757*	−0.1289**							0.4169
(Ancestry-adjusted)				(0.0393)	(0.0502)							(0.3328)
Ethnic diversity						0.4060	0.6235					−0.8059
						(0.5031)	(0.5154)					(0.5819)
Linguistic diversity								0.8101***	0.8436**			1.1046***
								(0.2990)	(0.3417)			(0.4017)
Religious diversity										1.0026**	0.3914	0.3020
										(0.4752)	(0.4113)	(0.3729)

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

TABLE 4. Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log − Likelihood value	−100,031.28	−118,022.86	−89,264.20	−117,706.09	−89,765.45	−120,938.37	−97,775.07	−115,146.47	−93,072.55	−110,347.06	−98,763.96	−81,199.39
Pseudo − R^2	0.90	0.88	0.91	0.88	0.91	0.88	0.90	0.88	0.91	0.89	0.90	0.92
$Corr[t_i, \widehat{t_i}]^2$	0.98	0.98	0.99	0.98	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.99
Number of countries	126	126	126	126	126	126	126	126	126	126	126	126
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Microgeographic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land productivity factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications between the years 1981 and 2016. Estimation methodology: Poisson regression model estimated by the Poisson maximum likelihood estimator. The exposure variable is the country's total population size averaged between the years 1981 and 2016. The expression $Corr[t_i, \widehat{t_i}]^2$ refers to the squared correlation between the observed (t_i) and predicted ($\widehat{t_i}$) number of scientific and technical publications. See the main text for additional details. ln $DRD4^{R2R7}$ refers to the log of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure (ethnicity-weighted). *Predicted Genetic Diversity* refers to the country's overall genetic diversity (expressed as expected heterozygosity) predicted by migratory distance from East Africa. *Migratory Distance from East Africa* refers to the distance from East Africa (i.e. Addis Ababa, Ethiopia (9N, 39E)) to the country's modern capital city (in 1000 km), restricting the migratory paths through five land-restricted way-points: namely Cairo, Egypt (30N, 31E); Istanbul, Turkey (41N, 28E); Phnom Penh, Cambodia (11N, 104E); Anadyr, Russia (64N, 177E); and Prince Rupert, Canada (54N, 130W). *Ethnic Diversity* refers to the country's ethnic diversity (0=low and 1=high). *Linguistic Diversity* refers to the country's linguistic diversity (0=low and 1=high). *Religious Diversity* refers to the country's religious diversity (0=low and 1=high). *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, America, Asia, and Europe. *Microgeographic Factors* include ln *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Area with Malaria*, *Population in Tropics*, and *Population in Temperate Zones*. Constant term included but not shown. Robust standard errors are reported between parenthesis.

*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

The estimates presented in columns (6) and (7) examine the sensitivity of the main results to the inclusion of ethnic diversity. The estimated regression coefficient associated with ethnic diversity enters with a positive sign in the regression model, but is not statistically significant at conventional significance levels. Again, the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure remains rather robust and precisely estimated.

In columns (8) to (11), I subsequently include measures of linguistic and religious diversity in the regression model. Interestingly, the results suggest that more linguistically diverse countries have on average a higher number of scientific and technical publications *per capita*, in the model specifications that also controls for the country-level $DRD4^{R2R7}$ measure. The positive impact of religious diversity on scientific knowledge creation is sensitive to the inclusion of the country-level $DRD4^{R2R7}$ variable. More importantly, the regression coefficient associated with the country-level $DRD4^{R2R7}$ variable remains highly robust to the issue of linguistic and religious diversity and statistically significant at the 1% significance level.

Finally, the results shown in column (12) test the sensitivity of the main findings to the simultaneous inclusion of the various diversity measures in the regression model.

Only the regression coefficient associated with predicted genetic and linguistic diversity remains statistically significant at conventional significance levels. Reassuringly, the main results regarding the country-level $DRD4^{R2R7}$ measure are unaffected even in this augmented model specification.

Including historical factors. In this section, I examine the sensitivity of the main findings to the inclusion of country-specific historical factors. The corresponding estimates are reported in Table 5. I first present the core findings, which include the full set of biogeographic and regional factors in the regression model, as shown in column (1). In the subsequent analysis, I examine the sensitivity of the regression coefficient associated with the country-level $DRD4^{R2R7}$ measure to the inclusion of various country-specific historical controls (e.g., colonial heritage, cultural background, and legal tradition), as shown in columns (2) to (7).

I start the empirical analysis by including the country's share of the European population in the regression equation, as shown in column (2). It has been argued that the kind of norms, values, and beliefs frequently attached to European populations might be particularly favorable to the process of economic development by facilitating the accumulation of human capital in society [Glaeser et al. (2004)]. Consistent with this idea, the estimated regression coefficient associated with the share of European population is positive and highly statistically significant at the 1% significance level. Even though the estimated regression coefficient associated with the country-level $DRD4^{R2R7}$ measure drops considerably in magnitude due to the inclusion of the share of European population in the regression model, it remains highly statistically significant at conventional significance levels.

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

TABLE 5. DRD4 exon III 2- and 7-repeat allele frequency and number of scientific and technical publications (Historical factors)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: Number of scientific and technical publications, t_i						
ln *DRD4*^{R2R7}	1.6474***	0.8628***	1.1629***	1.2884***	1.0275***	0.8963***	0.8984***
(Ethnicity-weighted)	(0.4456)	(0.3290)	(0.4016)	(0.3274)	(0.2492)	(0.3022)	(0.1962)
European population		3.5316***					1.9735***
(Share of total population)		(0.5820)					(0.6725)
Official language is English			0.5071**				0.0197
			(0.2075)				(0.2330)
Former colonizer is British				0.3197			−0.1659
				(0.3085)			(0.1707)
Former colonizer is French				−1.6669***			−0.9566**
				(0.4675)			(0.3720)
Former colonizer is Spanish				−1.4500***			−0.8357
				(0.5617)			(0.5630)
Former colonizer is Portuguese				−0.2805			0.3013
				(0.6406)			(0.6350)
Former colonizer is Dutch				−3.1787***			−3.1676***
				(0.4525)			(0.5223)
British legal origin					1.4919***		0.9639***
					(0.2389)		(0.3216)
French legal origin					0.7003***		0.5335***
					(0.2092)		(0.1487)
German legal origin					1.0072***		0.4584**
					(0.2234)		(0.2060)
Scandinavian legal origin					1.5993***		1.7173***
					(0.2635)		(0.3071)

TABLE 5. Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln protestants						0.1298***	−0.0090
						(0.0274)	(0.0302)
ln catholics						0.0803**	0.0960***
						(0.0350)	(0.0301)
ln muslims						0.0119	−0.0131
						(0.0269)	(0.0160)
Log − Likelihood value	−101,004.54	−68,376.14	−93,762.81	−63,904.01	−62,804.00	−74,230.16	−29,494.55
Pseudo − R^2	0.90	0.93	0.91	0.94	0.94	0.93	0.97
$Corr[t_i, \widehat{t_i}]^2$	0.98	0.99	0.98	0.99	0.99	0.99	1.00
Number of countries	128	128	128	128	128	128	128
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Microgeographic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land productivity factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications between the years 1981 and 2016. Estimation methodology: Poisson regression model estimated by the Poisson maximum likelihood estimator. The exposure variable is the country's total population size averaged between the years 1981 and 2016. The expression $Corr[t_i, \widehat{t_i}]^2$ refers to the squared correlation between the observed (t_i) and predicted $(\widehat{t_i})$ number of scientific and technical publications. See the main text for additional details. ln $DRD4^{R2R7}$ refers to the log of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure (ethnicity-weighted). *European Population* is the share of country's current population with European ancestry. *Official Language is English* refers to an indicator variable that takes a value of one if the country's official language is English and zero otherwise. *Former Colonizer Effects* refer to an indicator variable that takes a value of one if the country's former colonizer was British, French, Spanish, Portuguese, or Dutch, and zero otherwise. *Legal Origin Effects* refer to a set of indicator variables that takes a value of one if the country's legal tradition is British, French, German, or Scandinavian, and zero otherwise. *Major Religion Effects* refer to the log of the country's major religions (i.e. Protestants, Catholics, and Muslims) in % of the total population in the year 1980. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, America, Asia, and Europe. *Microgeographic Factors* include ln *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Area with Malaria*, *Population in Tropics*, and *Population in Temperate Zones*. Constant term included but not shown. Robust standard errors are reported between parenthesis.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

The next two model specifications assess the sensitivity of the baseline findings to the country's linguistic and colonial heritage, respectively. The estimates presented in column (3) show that the main findings are not confounded by a possible publication bias toward English-speaking journals. In addition, the main results presented in column (4) remain qualitatively unaffected by the inclusion of former major colonial power indicators (i.e. British, French, Spanish, Portuguese, or Dutch). The negative and statistically significant coefficient on the French, Spanish, and Dutch colonial indicators suggest a possible detrimental impact of colonization on the development of post-colonial countries with respect to scientific knowledge creation. Interestingly, the estimates provide no evidence that former British colonies are worse off with respect to contemporary scientific output, suggesting that the colonial practices (e.g., affecting the institutional, economic, and demographic settings) in these countries were quite different.

In column (5), I examine the importance of the legal tradition in the creation of scientific knowledge in society. Although all legal tradition indicators enter with a positive sign and are highly statistically significant at the 1% significance level, the estimates suggest that the largest impact on scientific output is found in countries with a British or Scandinavian legal tradition. The former estimate is consistent with the idea that a legal tradition that protects individual rights against arbitrary state interventions may facilitate freedom of expression, which appears quite important for the creation of knowledge in society. The positive and highly statistically significant coefficient associated with the Scandinavian legal tradition may reflect the high scientific productivity observed in Nordic countries (i.e. Denmark, Finland, Norway, and Sweden) triggered, for example, by high levels of R&D expenditures [OECD (2002); Sihvonen and Vähämaa (2015)]. Even though the regression coefficient of the country-level $DRD4^{R2R7}$ measure falls by 38% in this model specification, it remains highly statistically significant at the 1% significance level.

In column (6), I investigate the cultural hypothesis of scientific knowledge creation by including the log of the country's major religions (i.e. Protestants, Catholics, and Muslims) as percentages of the population in the regression model.

The estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure remains precisely estimated at the 1% significance level.

Finally, column (7) shows a "horse race" regression between the various historical controls. The results show that the country-level $DRD4^{R2R7}$ measure is robustly correlated with the number of scientific and technical publications *per capita*. With regard to the economic magnitude, the coefficient estimates in this final model specification suggest that increasing the country-level $DRD4^{R2R7}$ value of Poland (i.e. 0.1992) to the level of the United States (i.e. 0.2636) would, ceteris paribus, result in an increase of the number of scientific and technical journal articles *per capita* by about 25.16%. This change in the dependent variable is equivalent to moving Poland from the 77.34th percentile to the 78.91st percentile in the distribution of the number of scientific and technical journal articles *per capita*.

6. ROBUSTNESS ANALYSIS

In this section, I examine the robustness of the baseline estimates to the inclusion of technological frontier, as well as technological and human capability factors that might be important for the accumulation of scientific knowledge in society. In addition, I present results from the inclusion of various economic preference factors to rule out concerns regarding alternative channels in the relationship between knowledge creation and the country-level $DRD4^{R2R7}$ measure. I elaborate on these issues further below in the empirical analysis.

Technological frontier factors. In the following, I investigate the hypothesis that countries that are closer to the technological frontier (whether in genetic, linguistic, or religious terms) might find it easier to adopt advanced knowledge stocks and use them more efficiently for the accumulation of own scientific knowledge. The corresponding estimates are shown in Table 6. I consider three different entities that define the technological frontier at a global level: the United States, member countries of the OECD, and Neo-European countries. For each of these regions, I calculated the country's genetic, linguistic, and religious distance from the technological frontier, each capturing different aspects of human barriers to the diffusion of knowledge across countries. Again, I assessed the sensitivity of the estimated coefficient regarding the country-level $DRD4^{R2R7}$ measure relative to the inclusion of the various technological distance variables.

In columns (1) to (5), I present coefficient estimates for the country-level $DRD4^{R2R7}$ measure and the various technological distance variables relative to the United States. First of all, the main results regarding the key country-level $DRD4^{R2R7}$ measure remains qualitatively unaffected by the inclusion of these technological distance factors. Additionally, the estimates in column (5) suggest that genetic barriers to the United States as the technological frontier have a negative and statistically significant impact on scientific knowledge creation, consistent with the idea of cultural barriers to the diffusion of knowledge across countries [Spolaore and Wacziarg (2009)].

The results presented in columns (6) to (10) employ the member countries of the OECD as the global technological frontier. However, the main results remain qualitatively unaffected by this definition.

Finally, in model specifications (11) to (15), I define Neo-European countries (i.e. Australia, Canada, the United States, and New Zealand) as the global technological frontier. Again, the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure remain relatively robust and precisely estimated.

Technological and human capability factors. The creation of knowledge might depend on the country's technological and human resource base. Specifically, technologically advanced countries with a well-educated workforce might be quite efficient in the creation and absorption of scientific knowledge. To investigate this issue in more detail, I present coefficient estimates that control for a set of technological and human capability factors in the regression equation. It is worth mentioning that these controls might suffer from endogeneity problems.

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

TABLE 6. DRD4 exon III 2- and 7-repeat allele frequency and number of scientific and technical publications (Technological frontier factors)

	Technological frontier: USA					Technological frontier: OECD					Technological frontier: Neo-European				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Dependent variable: Number of scientific and technical publications, t_i														
ln *DRD4*^{R2R7} (Ethnicity-weighted)	1.6518***	1.3355***	1.2049***	1.5775***	1.3434***	1.6518***	1.8809***	1.0668***	1.5259***	1.4569***	1.6518***	1.6106***	1.0197**	1.3692***	1.1161***
	(0.4483)	(0.3778)	(0.4331)	(0.4653)	(0.3826)	(0.4483)	(0.3655)	(0.3931)	(0.4192)	(0.3248)	(0.4483)	(0.3735)	(0.4087)	(0.4087)	(0.3643)
Genetic distance to the frontier		−10.7393***			−12.6486***		−15.7422***			−13.2710***		−12.8205***			−10.4200***
		(3.1113)			(4.3840)		(3.6999)			(3.4877)		(3.9588)			(3.7209)
Linguistic distance to the frontier			−0.6183		−0.5348			−8.9403**		−5.0883			−1.2251*		−0.6624
			(0.5556)		(0.6307)			(3.7039)		(3.1796)			(0.6423)		(0.5815)
Religious distance to the frontier				−0.1963	1.1699				−2.3952	−1.2750				−3.5607*	−2.1111
				(0.7150)	(0.7595)				(1.5309)	(1.1350)				(1.9958)	(1.5960)

TABLE 6. Continued

	Technological frontier: USA					Technological frontier: OECD					Technological frontier: Neo-European				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Log – Likelihood value	−94,115.47	−81,357.46	−91,336.94	−93,927.52	−78,208.15	−94,115.47	−72,974.20	−86,131.13	−90,837.62	−69,918.05	−94,115.47	−78,660.85	−88,056.36	−89,300.89	−75,048.30
$Pseudo-R^2$	0.90	0.92	0.91	0.90	0.92	0.90	0.92	0.91	0.91	0.93	0.90	0.92	0.91	0.91	0.92
$Corr[t_i, \widehat{t_i}]^2$	0.98	0.99	0.98	0.98	0.99	0.98	0.99	0.98	0.98	0.99	0.98	0.99	0.98	0.98	0.99
Number of countries	126	126	126	126	126	126	126	126	126	126	126	126	126	126	126
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Microgeographic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land productivity factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications between the years 1981 and 2016. Estimation methodology: Poisson regression model estimated by the Poisson maximum likelihood estimator. The exposure variable is the country's total population size averaged between the years 1981 and 2016. The expression $Corr[t_i, \widehat{t_i}]^2$ refers to the squared correlation between the observed (t_i) and predicted ($\widehat{t_i}$) number of scientific and technical publications. See the main text for additional details. ln $DRD4^{R2R7}$ refers to the log of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure (ethnicity-weighted). *Genetic Distance to the Frontier* refers to the country's relative genetic distance to the technological frontier (i.e. U.S., OECD, and Neo-European countries). *Linguistic Distance to the Frontier* refers to the country's relative linguistic distance to the technological frontier (i.e. U.S., OECD, and Neo-European countries). *Religious Distance to the Frontier* refers to the country's relative religious distance to the technological frontier (i.e. U.S., OECD, and Neo-European countries). *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, America, Asia, and Europe. *Microgeographic Factors* include ln *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Area with Malaria*, *Population in Tropics*, and *Population in Temperate Zones*. Constant term included but not shown. Robust standard errors are reported between parenthesis.
*: Significant at the 10 level. **: Significant at the 5 level. ***: Significant at the 1 level.

Thus, the resulting regression coefficients should be interpreted with some caution.

However, the inclusion of these endogenous controls would not alter the main findings regarding the relationship between the country-level $DRD4^{R2R7}$ measure and scientific knowledge creation, since these are not the primary variables of interest but just additional country-level controls. The corresponding estimates are shown in Table 7. Notice that the baseline sample is reduced to 104 countries due to missing observations on the set of technological and human capability factors. Nevertheless, the main results remain qualitatively unaffected when restricting the baseline specification to the reduced 104-country sample, as shown in column (1).

First, I include the logarithm of GDP per capita in the regression model. The level of economic development can be interpreted as a summary measure capturing the standards of living in society (e.g., quality of public services, political institutions, and technology). The main results remain unaffected by the inclusion of GDP per capita in the regression model, as shown in column (2). In a similar vein, the estimates presented in column (3) show that the coefficient associated with the country-level $DRD4^{R2R7}$ measure is not sensitive to the inclusion of a human capital index. More importantly, the estimates in column (4) demonstrate very clearly that the level of resources devoted to R&D activities (measured as R&D expenditure over total GDP) is significantly correlated with the creation of scientific knowledge. Even though the regression coefficient associated with the key $DRD4^{R2R7}$ measure falls substantially in magnitude, its impact remains statistically different from zero at the 1% significance level. In columns (5) to (7), I examine the influence of infrastructure quality controls (measured as the number of internet users, fixed telephone, and mobile cellular subscriptions per 100 people) on the creation of scientific knowledge in society. All estimated coefficients associated with the various infrastructure quality measures are positive and statistically significant in all but one model specification. It is worth mentioning that internet access may also reflect openness to new ideas and people that appear beneficial for the creation of knowledge in society [Fagerberg and Srholec (2008)]. The regression coefficient associated with the country-level $DRD4^{R2R7}$ measure is virtually unaffected by the inclusion of infrastructure quality controls.

Finally, the estimates presented in column (8) assess the sensitivity of the main results to the simultaneous inclusion of the various technological and human capability factors in a single model specification. The relationship between scientific knowledge creation and the country-level $DRD4^{R2R7}$ measure remains positive and statistically significant at conventional significance levels.

Economic preference factors. As mentioned previously, the pleiotropic effects of DRD4 exon III polymorphism on various personality outcomes might undermine the interpretation of the country-level $DRD4^{R2R7}$ measure as full support for novelty-seeking behavior. Therefore, this section provides regression estimates that examine the sensitivity of the main findings to the inclusion of various

TABLE 7. DRD4 exon III 2- and 7-repeat allele frequency and number of scientific and technical publications (Technological and human capability factors)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: Number of scientific and technical publications, t_i							
ln $DRD4^{R2R7}$	1.5443***	1.2030***	1.3189***	1.0591***	0.5366**	0.7927***	1.5838***	0.6244***
(Ethnicity-weighted)	(0.4664)	(0.2427)	(0.3142)	(0.3642)	(0.2167)	(0.2325)	(0.4671)	(0.2005)
ln GDP per capita		1.0362***						0.6414***
		(0.1167)						(0.1713)
Human capital			0.7947***					0.1760
			(0.1884)					(0.1859)
R&D expenditures				0.5505***				0.1159
(Percentage of GDP)				(0.1144)				(0.1107)
Internet users					0.0642***			0.0260***
(Per 100 people)					(0.0070)			(0.0098)
Fixed telephone subscriptions						0.0418***		0.0049
(Per 100 people)						(0.0065)		(0.0070)
Mobile cellular subscriptions							0.0068	−0.0159***
(Per 100 people)							(0.0065)	(0.0045)

TABLE 7. Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log – Likelihood value	−90,032.54	−38,495.01	−70,424.21	−65,302.79	−37,952.14	−45,200.23	−88,422.64	−25,818.04
Pseudo – R^2	0.91	0.96	0.93	0.93	0.96	0.95	0.91	0.97
$Corr[t_i, \widehat{t_i}]^2$	0.98	0.99	0.99	0.98	0.99	0.99	0.98	1.00
Number of countries	104	104	104	104	104	104	104	104
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Microgeographic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land productivity factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications between the years 1981 and 2016. Estimation methodology: Poisson regression model estimated by the Poisson maximum likelihood estimator. The exposure variable is the country's total population size averaged between the years 1981 and 2016. The expression $Corr[t_i, \widehat{t_i}]^2$ refers to the squared correlation between the observed (t_i) and predicted $(\widehat{t_i})$ number of scientific and technical publications. See the main text for additional details. ln $DRD4^{R2R7}$ refers to the log of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure (ethnicity-weighted). *ln GDP per Capita* is the log of a country's GDP per capita averaged across the years 1981 to 2016. *Human Capital* is the country's human capital measure averaged across the years 1981 to 2016. *R&D Expenditures* is the country's flow of resources devoted to research and development activities (as percentage of total GDP) averaged across the years 1981 to 2016. *Internet Users* is the the number of internet users per 100 people averaged across the years 1981 to 2016. *Fixed Telephone Subscriptions* is the number of fixed telephone subscriptions per 100 people averaged across the years 1981 to 2016. *Mobile Cellular Subscriptions* is the number of mobile telephone cellular subscriptions per 100 people averaged across the years 1981 to 2016. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, America, Asia, and Europe. *Microgeographic Factors* include ln *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Area with Malaria*, *Population in Tropics*, and *Population in Temperate Zones*. Constant term included but not shown. Robust standard errors are reported between parenthesis.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

economic preference factors. Robustness of these results would rule out alternative explanations and at the same time provide suggestive evidence regarding the underlying mechanisms through which the DRD4 exon III polymorphism likely affects scientific knowledge creation.

The corresponding estimates are presented in Table 8. Due to the incomplete country coverage of the various economic preference factors in the study by Falk et al. (2018), the number of country observations falls to 71. However, replicating the baseline specification with this restricted sample size does not alter the main results substantially, as evidenced in column (1). Again, I assess the sensitivity of the country-level $DRD4^{R2R7}$ regression coefficient relative to the inclusion of the various economic preference factors. I begin by examining whether the main results are confounded by the economic preference factors that have attracted considerable attention in the literature regarding cross-country differences in economic development. Columns (2) and (4) provide evidence that patience and risk-taking preference factors, respectively, are strongly correlated with scientific knowledge creation in specifications without controlling for the country-level $DRD4^{R2R7}$ variable. The positive regression coefficients associated with both economic preference factors are consistent with the notion that patience and willingness to take risks are beneficial for the successful completion of uncertain and risky research activities. However, as columns (3) and (5) show, patience is the only preference factor that remains highly statistically significant once the country-level $DRD4^{R2R7}$ measure is accounted for.

The statistically not significant regression coefficient associated with the willingness to take risks once the regression model controls for the country-level $DRD4^{R2R7}$ variable in specification (5) suggests that DRD4 exon III polymorphism captures the extent of risk-taking in society to some degree. Accordingly, the drop in the regression coefficient of the $DRD4^{R2R7}$ variable in model specification (3) once adding patience to the model identifies another mechanism by which DRD4 exon III polymorphism affects scientific knowledge creation. Columns (6) to (13) document that the only significant factor of the remaining economic preference controls that is strongly correlated with scientific knowledge creation is negative reciprocity, as shown in columns (8) and (9). The positive regression coefficient associated with negative reciprocity may suggest support for the hypothesis that sanctioning of inefficient behaviors may help to sustain the kind of large-scale cooperation that facilitates scientific knowledge creation in society [Falk et al. (2018)]. Moreover, I find no evidence that the impact of the country-level $DRD4^{R2R7}$ measure is confounded by social preference measures, that is, positive reciprocity, altruism, or trust. Finally, column (14) presents results from a "horse race" specification between the various economic preference factors and the country-level $DRD4^{R2R7}$ measure. The estimates show that patience and negative reciprocity are the only economic preference factors that are significantly correlated with scientific knowledge creation in specifications that account for the country-level $DRD4^{R2R7}$ measure.

TABLE 8. DRD4 exon III 2- and 7-repeat allele frequency and number of scientific and technical publications (Economic preference factors)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Dependent variable: Number of scientific and technical publications, t_i													
ln $DRD4^{R2R7}$ (Ethnicity-weighted)	1.6695***		1.0843**		1.4862***		1.6941***		1.8349***		1.7098***		1.6705***	1.3649***
	(0.3498)		(0.4413)		(0.3952)		(0.3445)		(0.3296)		(0.3538)		(0.3656)	(0.4014)
Patience		1.2777***	0.7971**											1.2621***
		(0.2728)	(0.3367)											(0.4165)
Risk-taking				1.1021***	0.5147									−0.4293
				(0.3932)	(0.3601)									(0.3678)
Positive reciprocity						0.1267	−0.1468							0.2215
						(0.3846)	(0.2732)							(0.5579)
Negative reciprocity								0.5229*	0.8352***					0.8129***
								(0.3113)	(0.2445)					(0.2323)
Altruism										0.1725	−0.1470			−0.2841
										(0.3345)	(0.2191)			(0.2506)
Trust												0.5228	−0.0024	−0.6025
												(0.3516)	(0.2598)	(0.4633)

TABLE 8. Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Log − Likelihood value	−80,813.78	−81,629.52	−71,354.89	−104,834.53	−77,913.30	−119,893.07	−80,294.11	−115,611.63	−69,837.58	−119,423.13	−80,227.58	−113,368.96	−80,813.66	−57,837.31
$Pseudo - R^2$	0.91	0.91	0.92	0.88	0.91	0.87	0.91	0.87	0.92	0.87	0.91	0.87	0.91	0.94
$Corr[t_i, \widehat{t_i}]^2$	0.98	0.95	0.97	0.96	0.98	0.94	0.98	0.95	0.98	0.94	0.98	0.95	0.98	0.98
Number of countries	71	71	71	71	71	71	71	71	71	71	71	71	71	71
Microgeographic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land productivity factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications between the years 1981 and 2016. Estimation methodology: Poisson regression model estimated by the Poisson maximum likelihood estimator. The exposure variable is the country's total population size averaged between the years 1981 and 2016. The expression $Corr[t_i, \widehat{t_i}]^2$ refers to the squared correlation between the observed (t_i) and predicted $(\widehat{t_i})$ number of scientific and technical publications. See the main text for additional details. ln $DRD4^{R2R7}$ refers to the log of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure (ethnicity-weighted). *Patience* refers to the country's overall willingness to wait. *Risk-Taking* refers to the country's overall willingness to take risks in general. *Positive Reciprocity* refers to the country's overall willingness to return a gift for favorable help. *Negative Reciprocity* refers to the country's overall willingness to take revenge for unfair behavior. *Altruism* refers to the country's overall willingness to donate for good causes. *Trust* refers to the country's overall perception that other people only have the best intentions. *Microgeographic Factors* include ln *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Area with Malaria*, *Population in Tropics*, and *Population in Temperate Zones*. Constant term included but not shown. Robust standard errors are reported between parenthesis.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

TABLE 9. DRD4 exon III 2- and 7-repeat allele frequency and number of scientific and technical publications (Extended sample size)

	(1) Table 3 Column (7)	(2) Table 4 Column (12)	(3) Table 5 Column (7)	(4) Table 6 Column (5)	(5) Table 6 Column (10)	(6) Table 6 Column (15)	(7) Table 7 Column (8)
	Dependent variable: Number of scientific and technical publications, t_i						
ln $DRD4^{R2R7}$	1.7362***	1.7469***	1.0010***	1.4760***	1.8359***	1.2167***	0.5860***
(Ethnicity-weighted)	(0.3214)	(0.3151)	(0.2165)	(0.3849)	(0.3950)	(0.4114)	(0.1981)
ln Absolute latitude	0.8979**	0.7114**	−0.2377	0.5540	0.5544	0.5096	0.3006
	(0.3737)	(0.3550)	(0.1771)	(0.3655)	(0.3662)	(0.3616)	(0.2443)
Percentage arable land area	0.0164**	0.0076	−0.0001	0.0045	0.0120	0.0044	0.0156***
	(0.0072)	(0.0085)	(0.0073)	(0.0089)	(0.0079)	(0.0082)	(0.0050)
Agricultural suitability	−1.9420***	−1.6137***	−0.6588	−1.1919**	−1.4419***	−1.0447*	−0.9951***
	(0.5443)	(0.5331)	(0.5436)	(0.5437)	(0.5219)	(0.5383)	(0.3064)
Terrain roughness	0.3520	0.6118	0.0700	0.2713	0.3296	0.2385	2.0551***
	(0.8798)	(0.7551)	(0.7299)	(0.7550)	(0.8175)	(0.8180)	(0.4914)
Temperature	0.0653***	0.0540***	0.0055	0.0414*	0.0408**	0.0408**	0.0187
	(0.0201)	(0.0163)	(0.0110)	(0.0221)	(0.0171)	(0.0192)	(0.0131)
Precipitation	0.0074**	0.0141***	0.0039	0.0091**	0.0085**	0.0071**	−0.0035*
	(0.0035)	(0.0040)	(0.0031)	(0.0039)	(0.0036)	(0.0036)	(0.0021)
Population in area with malaria	−2.0843***	−2.6162***	−1.9286***	−2.3970***	−2.4798***	−2.1746***	−1.6820***
(Share of total Population)	(0.5753)	(0.6862)	(0.5161)	(0.6059)	(0.6601)	(0.6235)	(0.3354)
Population in Tropics	−0.6168	−1.6852**	−1.6487***	−1.1750	−1.1390	−0.9483	0.9829**
(Share of total population)	(0.6865)	(0.8122)	(0.5616)	(0.8412)	(0.8488)	(0.8323)	(0.4284)
Population in temperate zones	2.8561***	2.2774***	1.1355	2.2808***	2.6405***	2.3989***	0.6144**
(Share of total population)	(0.8215)	(0.6915)	(0.7378)	(0.8291)	(0.7988)	(0.8131)	(0.2431)

TABLE 9. Continued

	(1) Table 3 Column (7)	(2) Table 4 Column (12)	(3) Table 5 Column (7)	(4) Table 6 Column (5)	(5) Table 6 Column (10)	(6) Table 6 Column (15)	(7) Table 7 Column (8)
Log − Likelihood value	−111,865.62	−89,994.61	−38,085.65	−90,759.80	−93,070.71	−91,584.08	−29,204.37
Pseudo $-R^2$	0.89	0.91	0.96	0.91	0.90	0.90	0.97
$Corr\left[t_i, \widehat{t_i}\right]^2$	0.98	0.99	1.00	0.98	0.98	0.98	0.99
Number of countries	140	133	135	130	130	130	110
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diversity factors	No	Yes	No	No	No	No	No
Historical factors	No	No	Yes	No	No	No	No
Technological frontier factors	No	No	No	U.S.	OECD	Neo-European	No
Technological and human capability factors	No	No	No	No	No	No	Yes

Notes: The dependent variable is the average number of scientific and technical publications between the years 1981 and 2016. Estimation methodology: Poisson regression model estimated by the Poisson maximum likelihood estimator. The exposure variable is the country's total population size averaged between the years 1981 and 2016. The expression $Corr[t_i, \widehat{t_i}]^2$ refers to the squared correlation between the observed (t_i) and predicted $(\widehat{t_i})$ number of scientific and technical publications. See the main text for additional details. ln $DRD4^{R2R7}$ refers to the log of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure (ethnicity-weighted). ln *Absolute Latitude* is the log of the absolute value of a country's approximate centroid latitude in decimal degrees. *Percentage Arable Land Area* is the country's percentage arable land area. *Agricultural Suitability* is a geospatial indicator, ranging from 0 to 1, of land suitability for agriculture across 0.5 decimal degrees latitude × longitude grid cells in the area covered by each country. *Terrain Roughness* is a geospatial indicator that indicates the average absolute change in elevation values across contiguous 1 decimal degrees latitude × longitude grid cells in the area covered by each country. *Temperature* is the mean country's temperature (in degree celsius) during the period 1960 and 1990. *Precipitation* is the mean country's precipitation (in total millimeters per month) during the period 1960 and 1990. *Population in Area with Malaria* is the share of a country's population in 1995 residing in areas contracting with malaria falciparum. *Population in Tropics* is the share of a country's population in 1995 residing in tropics. *Population in Temperate Zones* is the share of a country's population in 1995 residing in temperate zones. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, America, Asia, and Europe. *Diversity Factors* include *Predicted Genetic Diversity*, *Migratory Distance from East Africa*, *Ethnic Diversity*, and *Linguistic Diversity*. *Historical Factors* include *European Population Share*, *Official Language is English*, *Former Colonizer Effects*, *Legal Origin Effects*, and *Major Religion Effects*. *Technological Frontier Factors* include *Genetic Distance to the Frontier*, *Linguistic Distance to the Frontier*, and *Religious Distance to the Frontier*. *Technological and Human Capability Factors* include *ln GDP per Capita*, *Human Capital*, *R&D Expenditures*, *Internet Users*, *Fixed Telephone Subscriptions*, and *Mobile Cellular Subscriptions*. Constant term included but not shown. Robust standard errors are reported between parenthesis.

*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

TABLE 10. DRD4 exon III 2- and 7-repeat allele frequency and number of scientific and technical publications (Robustness to ethnic diversity)

	(1) Table 3 Column (7)	(2) Table 5 Column (7)	(3) Table 6 Column (5)	(4) Table 6 Column (10)	(5) Table 6 Column (15)	(6) Table 7 Column (8)
	Dependent variable: Number of scientific and technical publications, t_i					
ln $DRD4^{R2R7}$	1.6922***	0.9028***	1.3423***	1.4572***	1.1274***	0.6422***
(Ethnicity-weighted)	(0.4134)	(0.1948)	(0.3699)	(0.3236)	(0.3596)	(0.1956)
Ethnic diversity	0.6591	0.0676	0.3301	0.0340	0.1872	0.3880
	(0.5173)	(0.2740)	(0.4066)	(0.3888)	(0.4032)	(0.2791)
$Log-Likelihood\ value$	−98,520.62	−29,477.92	−77,620.60	−69,912.05	−74,862.13	−25,278.68
$Pseudo-R^2$	0.90	0.97	0.92	0.93	0.92	0.97
$Corr[t_i,\widehat{t_i}]^2$	0.98	1.00	0.99	0.99	0.99	1.00
Number of countries	129	128	126	126	126	104

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

Table 10. Continued

	(1) Table 3 Column (7)	(2) Table 5 Column (7)	(3) Table 6 Column (5)	(4) Table 6 Column (10)	(5) Table 6 Column (15)	(6) Table 7 Column (8)
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Microgeographic factors	Yes	Yes	Yes	Yes	Yes	Yes
Land productivity factors	Yes	Yes	Yes	Yes	Yes	Yes
Climatic factors	Yes	Yes	Yes	Yes	Yes	Yes
Health factors	Yes	Yes	Yes	Yes	Yes	Yes
Historical factors	No	Yes	No	No	No	No
Technological frontier factors	No	No	U.S.	OECD	Neo-European	No
Technological and human capability factors	No	No	No	No	No	Yes

Notes: The dependent variable is the average number of scientific and technical publications between the years 1981 and 2016. Estimation methodology: Poisson regression model estimated by the Poisson maximum likelihood estimator. The exposure variable is the country's total population size averaged between the years 1981 and 2016. The expression $Corr[t_i, \widehat{t_i}]^2$ refers to the squared correlation between the observed (t_i) and predicted ($\widehat{t_i}$) number of scientific and technical publications. See the main text for additional details. ln $DRD4^{R2R7}$ refers to the log of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure (ethnicity-weighted). *Ethnic Diversity* refers to the country's ethnic diversity (0=low and 1=high). *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, America, Asia, and Europe. *Microgeographic Factors* include ln *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Area with Malaria*, *Population in Tropics*, and *Population in Temperate Zones*. *Historical Factors* include *European Population* share, *Official Language is English*, *Former Colonizer Effects*, *Legal Origin Effects*, and *Major Religion Effects*. *Technological Frontier Factors* include *Genetic Distance to the Frontier*, *Linguistic Distance to the Frontier*, and *Religious Distance to the Frontier*. *Technological and Human Capability Factors* include *ln GDP per Capita*, *Human Capital*, *R&D Expenditures*, *Internet Users*, *Fixed Telephone Subscriptions*, and *Mobile Cellular Subscriptions*. Constant term included but not shown. Robust standard errors are reported between parenthesis.

*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

Additional robustness tests. In this section, I examine the robustness of the main findings to alternative model specifications. In Table 9, I provide regression results of simpler model specifications. In particular, in order to investigate whether the regression analysis holds for a larger set of countries, even with fewer country-level controls, I drop the variables *ln Distance to Major Markets*, *ln Distance to Coast or River*, and *ln Hydrocarbons per Person* from the baseline specifications, since these are the variables that significantly restrict the number of countries in the estimation sample. Replicating the main analysis without these microgeographic controls does not significantly affect the main results regarding the positive impact of the country-level $DRD4^{R2R7}$ measure on scientific knowledge creation.

Since the country-level $DRD4^{R2R7}$ measure is constructed using information on the ethnic composition of each country, it could be that the main results partly reflect the impact of ethnic diversity on scientific knowledge creation. To rule out this concern, the results presented in Table 10 replicates the main analysis by including ethnic diversity into the regression model. Overall, the regression coefficient associated with the country-level $DRD4^{R2R7}$ measure remains highly robust and precisely estimated throughout all model specifications, while the regression coefficient of ethnic diversity is statistically not significant at conventional significance levels.

7. CONCLUSION

This paper establishes the beneficial effects of novelty-seeking traits on the level of scientific knowledge creation in society. The key explanatory variable refers to the human DRD4 exon III 2- and 7-repeat allele variants that candidate gene studies of personality have linked to the human phenotype of novelty-seeking behavior. The issue of novelty-seeking has been linked to beneficial economic attitudes related to risk-taking, creativity, and entrepreneurship that constitute the primary source of technological progress in Schumpeterian-inspired endogenous growth models.

The baseline estimates suggest that a one standard deviation increase in the prevalence of novelty-seeking traits would, ceteris paribus, increase the number of scientific and technical journal articles *per capita* by about 25.16%. This result is robust to the inclusion of potentially confounding microgeographic, land productivity, climatic, health, legal, cultural, colonial, and regional factors. The estimated magnitude is substantial, as it is equivalent to moving Poland from the 77.34th percentile to the 78.91st percentile in the distribution of the number of scientific and technical journal articles *per capita*. The main findings remain qualitatively robust to the inclusion of additional determinants in cross-country differences of scientific knowledge creation (e.g., cultural distance to the technological frontier and the country's technological and human resource base) and to the inclusion of various economic preference factors to rule out concerns of alternative

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

mechanisms regarding the interpretation of DRD4 exon III polymorphism in the empirical analysis.

This research contributes significantly to our understanding regarding the importance of some deep-rooted historical factors in the creation of scientific knowledge in society. The findings should not be misinterpreted to suggest that the prevalence of novelty-seeking traits is the only factor explaining differences in scientific knowledge creation across countries. Rather, the main findings show that after controlling for a large set of technological, biogeographical, cultural, and institutional determinants, there remains a positive and statistically significant effect of the country-level $DRD4^{R2R7}$ measure on the number of scientific and technical journal articles *per capita*. This observation highlights the importance of socioeconomic factors that might help to facilitate the creation and diffusion of knowledge across societies.

NOTES

1. For example, Romer (1990) illustrates that product innovations such as magnetic tapes and home videocassette recorders would not have been possible without academic research in the field of electromagnetism.

2. Strictly speaking, the probability that an outside firm will produce exactly one innovation during the length of time $d\tau$ is, according to the Poisson distribution in equation (14), given by $f_T[T=1 \mid \lambda(\psi, \ell_t^R, R, z, \nu)] = \lambda(\psi, \ell_t^R, R, z, \nu)d\tau e^{-\lambda(\psi, \ell_t^R, R, z, \nu)d\tau}$. Dividing both sides by $d\tau$ and taking the limit $d\tau \to 0$ yields $\lim_{d\tau\to 0}(\frac{1}{d\tau} f_T[T=1 \mid \lambda(\psi, \ell_t^R, R, z, \nu)]) = \lambda(\psi, \ell_t^R, R, z, \nu)$, which approximately equals the probability of one innovation during an infinitesimally small time duration.

3. Another possibility would be to model the arrival rate to follow a sigmoid function (e.g., the cumulative distribution function of the logistic function) to account for threshold effects in the research sector [Aghion and Howitt (1998)].

4. To be more precise, $\boldsymbol{x}_i' = \left(\ln(1-\psi_i), \ln R_i, \ln z_i, \ln(\frac{1-\alpha}{\alpha}), \ln[\frac{\gamma_i}{1+\gamma_i(\frac{1-\alpha}{\alpha})}]\right)$ is a (1×5) vector of explanatory variables and $\boldsymbol{\theta} = (1, \varphi, (1-\beta), 1, 1)'$ is the corresponding (5×1) parameter vector.

5. The interested reader is referred to Cameron and Trivedi (2013) for an excellent introduction to model specification and estimation of count data applications.

6. The interested reader is referred to the supplemental Appendix A for additional details on data construction and sources of the various control variables. Descriptive statistics for the various estimation samples employed in the empirical analysis are provided in supplemental Appendix B.

7. See Ashraf and Galor (2013) for a similar approach to approximate the stock of knowledge in the aggregate economy.

8. In biology, genetic information in every cell is present in the form of chromosomes, which are made up of long threads of relatively small nucleotides. These long threads are the carriers of genetic information and are called DNA. A particular gene corresponds to a shorter segment of the DNA thread along a specific chromosome. The genes are actually responsible for specific activities in the cells. Each gene can take many different forms, which are called alleles of that gene. For example, the specific gene that is responsible for different blood types contains different allele variants (A, B, and O). In molecular genetics, a gene is called *polymorphic* if population geneticists detected more than two different allele variants in populations. The interested reader is referred to Cavalli-Sforza et al. (1994) for a more detailed discussion of the definitions and methods employed in molecular genetics.

9. In the empirical analysis, I control for these alternative channels and their possible impact on the point estimates associated with the country-level DRD4 exon III 2- and 7-repeat allele frequency measure. I would like to thank an anonymous referee for suggesting this important point in the empirical analysis.

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

10. The interested reader is referred to Gören (2016) for a detailed discussion of the sampled populations and corresponding data sources.

11. I use data on mean total population size from the *Penn World Table, version 9.1* averaged between the years 1981 and 2016 [Feenstra et al. (2015)].

12. This difference in the $DRD4^{R2R7}$ measure between Poland and the United States would roughly correspond to a one standard deviation change in the country-level $DRD4^{R2R7}$ measure in the 129-country sample.

REFERENCES

Acemoglu, D., S. Johnson and J. A. Robinson (2001) The colonial origins of comparative development: An empirical investigation. *The American Economic Review* 91, 1369–1401.

Acemoglu, D., S. Johnson and J. A. Robinson (2002) Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly Journal of Economics* 117, 1231–1294.

Aghion, P. and P. Howitt (1992) A model of growth through creative destruction. *Econometrica* 60, 323–351.

Aghion, P. and P. Howitt (1998) *Endogenous Growth Theory*. Cambridge, Massachusetts: The MIT Press.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat and R. Wacziarg (2003) Fractionalization. *Journal of Economic Growth* 8, 155–194.

Algan, Y. and P. Cahuc (2013) Trust and growth. *Annual Review of Economics* 5(1), 521–549.

Arbatli, C. E., Q. H. Ashraf, O. Galor, and M. Klemp (2020) Diversity and conflict. *Econometrica* 88, 727–797.

Asghari, V., S. Sanyal, S. Buchwaldt, A. Paterson, V. Jovanovic and H. H. M. Van Tol (1995) Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants. *Journal of Neurochemistry* 65, 1157–1165.

Ashraf, Q. and O. Galor (2013) The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic development. *The American Economic Review* 103, 1–46.

Basu, S. and D. N. Weil (1998) Appropriate technology and growth. *The Quarterly Journal of Economics* 113, 1025–1054.

Benjamin, J., L. Li, C. Patterson, B. D. Greenberg, D. L. Murphy and D. H. Hamer (1996) Population and familial association between the D4 dopamine receptor gene and measures of novelty seeking. *Nature* 12, 81–84.

Bobb, A. J., F. X. Castellanos, A. M. Addington and J. L. Rapoport (2006) Molecular genetic studies of ADHD: 1991 to 2004. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* 141B, 551–565.

Bonin, H., T. Dohmen, A. Falk, D. Huffman and U. Sunde (2007) Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics* 14(6), 926–937.

Cameron, A. C. and P. K. Trivedi (2013) *Regression Analysis of Count Data*. New York, , 2nd ed. USA: Cambridge University Press.

Cann, H. M., C. de Toma, L. Cazes, M. F. Legrand, V. M. L. P. J. Bodmer, W. F. Bodmer, B. Bonne-Tamir, A. Cambon-Thomsen, Z. Chen, J. Chu, C. Carcassi, L. Contu, R. Du, L. Excoffier, G. B. Ferrara, J. Friedlaender, H. Groot, D. Gurwitz, T. Jenkins, R. J. Herrera, X. H. J. Kidd, K. K. Kidd, A. Langaney, A. A. Lin, S. Q. Mehdi, P. Parham, A. Piazza, M. P. Pistillo, Y. Qian, Q. Shu, J. Xu, S. Zhu, J. L. Weber, H. T. Greely, M. W. Feldman, G. Thomas, J. Dausset, and L. L. Cavalli-Sforza (2002) A human genome diversity cell line panel. *Science* 296, 261–262.

Carpenter, J. P., J. R. Garcia and J. K. Lum (2011) Dopamine receptor genes predict risk preferences, time preferences, and related economic choices. *Journal of Risk and Uncertainty* 42, 233–261.

Cavalli-Sforza, L. L., P. Menozzi and A. Piazza (1994) *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press

Chabris, C. F., B. M. Hebert, D. J. Benjamin, J. Beauchamp, D. Cesarini, M. van der Loos, M. Johannesson, P. K. E. Magnusson, P. Lichtenstein, C. S. Atwood, J. Freese, T. S. Hauser, R. M. Hauser, N. Christakis and D. Laibson (2012) Most reported genetic associations with general intelligence are probably false positives. *Psychological Science* 23, 1314–1323.

Chen, C., M. Burton, E. Greenberger and J. Dmitrieva (1999) Population migration and the variation of Dopamine D4 Receptor (DRD4) allele frequencies around the globe. *Evolution and Human Behavior* 20, 309–324.

Civelli, O., J. R. Bunzow, D. K. Grandy, Q. Y. Zhou and H. H. M. Van Tol (1991) Molecular biology of the dopamine receptors. *European Journal of Pharmacology: Molecular Pharmacology* 207, 277–286.

Cloninger, C. R. (1987) A systematic method for clinical description and classification of personality variants. A proposal. *Archives of General Psychiatry* 44, 573–588.

Cochran, G. and H. Harpending (2009) *The 10,000 Year Explosion: How Civilization Accelerated Human Evolution*. New York: Basic Books.

Cohen, W. M. and D. A. Levinthal (1990) Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35, 128–152.

Cook, C. J. (2014) The role of lactase persistence in precolonial development. *Journal of Economic Growth* 19, 369–406.

Cook, C. J. (2015) The natural selection of infectious disease resistance and its effect on contemporary health. *The Review of Economics and Statistics* 97, 742–757.

Dannemann, B. C. and E. Gören (2018) The educational burden of ADHD: Evidence from student achievement test scores. *University of Oldenburg, Department of Economics, Working Paper No. V-408-18*.

Desmet, K., I. Ortuño-Ortín and R. Wacziarg (2012) The political economy of linguistic cleavages. *Journal of Development Economics* 97, 322–338.

Dohmen, T., A. Falk, D. Huffman and U. Sunde (2010) Are risk aversion and impatience related to cognitive ability? *American Economic Review* 100(3), 1238–1260.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp and G. G. Wagner (2011) Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association* 9(3), 522–550.

Dreber, A., C. L. Apicella, D. T. A. Eisenberg, J. R. Garcia, R. S. Zamore, J. K. Lum and B. Campbell (2009) The 7R polymorphism in the dopamine receptor D_4 gene (DRD4) is associated with financial risk taking in men. *Evolution and Human Behavior* 30, 85–92.

Easterly, W. and R. Levine (2016) The european origins of economic development. *Journal of Economic Growth* 21, 225–257.

Ebstein, R. P. (1997) Saga of an adventure gene: Novelty seeking, substance abuse and the Dopamine D4 Receptor (D4DR) Exon III repeat polymorphism. *Molecular Psychiatry* 2, 381–384.

Ebstein, R. P., O. Novick, R. Umansky, B. Priel, Y. Osher, D. Blaine, E. R. Bennett, L. Nemanov, M. Katz and R. H. Belmaker (1996) Dopamine D4 Receptor (D4DR) Exon III polymorphism associated with the human personality trait of novelty seeking. *Nature* 12, 78–80.

Epper, T., E. Fehr, H. Fehr-Duda, C. T. Kreiner, D. D. Lassen, S. Leth-Petersen and G. N. Rasmussen (2020) Time discounting and wealth inequality. *American Economic Review* 110(4), 1177–1205.

Fagerberg, J. and M. Srholec (2008) National innovation systems, capabilities and economic development. *Research Policy* 37, 1417–1435.

Fagerberg, J., M. Srholec and B. Verspagen (2010) Innovation and economic development. In K. J. Arrow and M. D. Intriligator, eds., *Handbook of the Economics of Innovation*, Chapter 20, pp. 833–872. Elsevier.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman and U. Sunde (2018) Global evidence on economic preferences. *The Quarterly Journal of Economics* 133(4), 1645–1692.

Faraone, S. V., C. Bonvicini and C. Scassellati (2014) Biomarkers in the diagnosis of ADHD - Promising directions. *Current Psychiatry Reports* 16, 497.

Faraone, S. V., R. H. Perlis, A. E. Doyle, J. W. Smoller, J. J. Goralnick, M. A. Holmgren and P. Sklar (2005) Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry* 57, 1313–1323.

Feenstra, R. C., R. Inklaar and M. P. Timmer (2015) The next generation of the penn world table. *American Economic Review* 105, 3150–3182.

Gallup, J. L., J. D. Sachs and A. D. Mellinger (1999) Geography and economic development. *International Regional Science Review* 2, 179–232.

Galor, O. and S. Michalopoulos (2012) Evolution and the growth process: Natural selection of entrepreneurial traits. *Journal of Economic Theory* 147, 759–780.

Gizer, I. R., C. Ficks and I. D. Waldman (2009) Candidate gene studies of ADHD: A meta-analytical review. *Human Genetics* 126, 51–90.

Glaeser, E. L., R. L. Porta, F. Lopez-De-Silanes and A. Shleifer (2004) Do institutions cause growth? *Journal of Economic Growth*, 9, 271–303.

Global Mapping International (2010) World Language Mapping System, Version 16. http://worldgeodatasets.com/language/index.html.

Gören, E. (2014) How ethnic diversity affects economic growth. *World Development* 59, 275–297.

Gören, E. (2016) The biogeographic origins of novelty-seeking traits. *Evolution and Human Behavior* 37, 456–469.

Gören, E. (2017) The persistent effects of novelty-seeking traits on comparative economic development. *Journal of Development Economics* 126, 112–126.

Gorodnichenko, Y. and G. Roland (2017) Culture, institutions and the wealth of nations. *The Review of Economics and Statistics* 99, 402–416.

Grossman, G. M. and E. Helpman (1991) Quality ladders in the theory of growth. *The Review of Economic Studies* 58, 43–61.

Hong, L. and S. E. Page (2001) Problem solving by heterogenous agents. *Journal of Economic Theory* 91, 123–163.

Hong, L. and S. E. Page (2004) Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences* 101, 16385–16389.

Jensen, P., D. Mrazek, P. K. Knapp, L. Steinberg, C. Pfeffer, J. Schowalter and T. Shapiro (1997) Evolution and revolution in child psychiatry: ADHD as a disorder of adaptation. *Journal of the American Academy of Child and Adolescent Psychiatry* 36, 1672–1679.

Jiang, Y., S. H. Chew and R. P. Ebstein (2013) The role of D4 receptor gene exon III polymorphisms in shaping human altruism and prosocial behavior. *Frontiers in Human Neuroscience* 7, 1–7.

Kerr, W. (2008) Ethnic scientific communities and international knowledge diffusion. *The Review of Economics and Statistics* 90, 518–537.

Knack, S. and Keefer, P. (1997) Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics* 112(4), 1251–1288.

Kuhnen, C. M. and J. Y. Chiao (2009) Genetic determinants of financial risk taking. *PLoS ONE* 4, 1–4.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1999) The quality of government. *The Journal of Law, Economics, & Organization* 15, 222–279.

LaHoste, G. J., J. M. Swanson, S. B. Wigal, C. Glabe, T. Wigal, N. King and J. L. Kennedy (1996) Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder. *Molecular Psychiatry* 1, 121–124.

Lichter, J. B., C. L. Barr, J. L. Kennedy, H. H. V. Tol, K. K. Kidd and K. J. Livak (1993) A hypervariable segment in the human dopamine receptor D_4 (DRD4) gene. *Human Molecular Genetics* 2, 767–773.

Mayseless, N., F. Uzefovsky, I. Shalev, R. P. Ebstein and S. G. Shamay-Tsoory (2013) The association between creativity and 7R polymorphism in the dopamine receptor D4 gene (DRD4). *Frontiers in Human Neuroscience* 7, 1–7.

Oak, J. N., J. Oldenhof and H. H. V. Tol (2000) The dopamine D_4 receptor: One decade of research. *European Journal of Pharmacology* 405, 303–327.

OECD (2002) OECD Science, Technology, and Industry Outlook 2002. *OECD Publications Service*. http://www.oecd-ilibrary.org/industry-and-services/oecd-science-technology-and-industry-outlook-2002_sti_outlook-2002-en.

OECD (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities*. OECD Publishing. http://www.oecd-ilibrary.org/ science-and-technology/frascati-manual-2015_9789264239012-en, Paris.

Paterson, A. D., G. A. Sunohara and J. L. Kennedy (1999) Dopamine D4 receptor gene: Novelty or nonsense? *Neuropsychopharmacology* 21, 3–16.

Plomin, R., J. C. DeFries, V. S. Knopik and J. M. Neiderhiser (2013) *Behavioral Genetics*. New York: Worth Publishers.

Putterman, L. and D. N. Weil (2010) Post-1500 population flows and the long run determinants of economic growth and inequality. *The Quarterly Journal of Economics* 125, 1627–1682.

Romer, P. M. (1990) Endogenous technological change. *Journal of Political Economy* 98, 71–102.

Sachs, J. D. and A. M. Warner (2001) The curse of natural resources. *European Economic Review* 45, 827–838.

Sihvonen, J. and S. Vähämaa (2015) Business research in the nordic countries: An analysis of research output across countries, disciplines, and institutions. *Nordic Journal of Business* 64, 266–296.

Spolaore, E. and R. Wacziarg (2009) The diffusion of development. *The Quarterly Journal of Economics* 124, 469–529.

Spolaore, E. and R. Wacziarg (2012) The long-term barriers to the international diffusion of innovations. *National Bureau of Economic Research*, 11–46.

Sutter, M., M. G. Kocher, D. Glätzle-Rützler and S. T. Trautmann (2013) Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review* 103(1), 510–531.

Swanson, J., J. Oosterlaan, M. Murias, S. Schuck, P. Flodman, M. A. Spence, M. Wasdell, Y. Ding, H.-C. Chi, M. Smith, M. Mann, C. Carlson, J. L. Kennedy, J. A. Sergeant, P. Leung, Y.-P. Zhangi, A. Sadeh, C. Chen, C. K. Whalen, K. A. Babb, R. Moyzis and M. I. Posner (2000) Attention deficit/hyperactivity disorder children with a 7-repeat allele of the dopamine receptor D4 gene have extreme behavior but normal performance on critical neuropsychological tests of attention. *Proceedings of the National Academy of Sciences* 97, 4754–4759.

Van Tol, H. H. M., J. R. Bunzow, H.-C. Guan, R. K. Sunahara, P. Seeman, H. B. Niznik and O. Civelli (1991) Cloning of the gene for a human dopamine D_4 receptor with high affinity for the antipsychotic clozapine. *Nature* 350, 610–614.

Van Tol, H. H. M., C. M. Wu, H.-C. Guan, K. Ohara, J. R. Bunzow, O. Civelli, J. Kennedy, P. Seeman, H. B. Niznik and V. Jovanovic (1992) Multiple dopamine D4 receptor variants in the human population. *Nature* 358, 149–152.

Viscusi, W. K. and J. Hersch (2001) Cigarette smokers as job risk takers. *Review of Economics and Statistics* 83(2), 269–280.

Wang, E., Y.-C. Ding, P. Flodman, J. R. Kidd, K. K. Kidd, D. L. Grady, O. A. Ryder, M. A. Spence, J. M. Swanson and R. K. Moyzis (2004) The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. *American Journal of Human Genetics* 74, 931–944.

Weber, M. (1958) *The Protestant Ethic and the Spirit of Capitalism*. New York: Charles Scribner's Sons.

Wu, J., H. Xiao, H. Sun, L. Zou, and L.-Q. Zhu (2012) Role of dopamine receptors in ADHD: A systematic meta-analysis. *Molecular Neurobiology* 45, 605–620.

https://doi.org/10.1017/S1365100521000092 Published online by Cambridge University Press