Finding Young Einsteins: Olympiads and STEM Talent Discovery

Ruchir Agarwal, Patrick Gaule and Yuyan Jiang*

September 11, 2024

Abstract

Discovering and supporting young talent early can speed up progress in science and technology. Yet, there is little systematic evidence on how to discover promising STEM talent at an early stage. In this paper, we explore whether the International Science Olympiads – competitions for top high school students – can be used for STEM talent discovery. Using novel hand-collected data, we present three main findings: (1) Among the top 500 Science Olympiad scorers annually, over 50% achieve PhDs, far exceeding the US average of 1% and the global average of 0.2%. (2) Among major scientific award winners, 5% have taken part in Science Olympiads, rising to 13% for mathematics awards. (3) International Mathematical Olympiad (IMO) participants excel in both math and other fields – securing 80% of the scientific awards that are awarded to Olympiad alumni over time. Moreover, IMO medalists become scientific prize winners at thirty times the rate of undergraduate alumni from the world's top ten universities (seventy-five times for IMO gold medalists). Overall, our results suggest that high school competitions, and the IMO in particular, can be an excellent way to discover promising STEM talent at an early stage.

Keywords: Talent, Innovation, Universities, Scientific Knowledge Production

*Agarwal: Harvard University and Global Talent Lab RAgarwal@imf.org. Gaule: University of Bristol, patrick.gaule @bristol.ac.uk. Jiang: University of Cambridge. Acknowledgements: We acknowledge financial support from Founders' Pledge, Polynera and Andrew Shaffer. We thank Archana Tiwari and Nisha Binu for research assistance, Deivis Angeli for comments, and Chris Peterson for his insights on talent discovery. All errors are our own.

1 Introduction

Talented individuals play a vital role in advancing science and technology. Historically, policies in science and innovation have mainly targeted individuals already pursuing scientific careers, offering them grants, fellowships, and awards to boost their research. However, less emphasis has been placed on the early identification and nurturing of potential innovators. Can discovering and supporting talented youth early on accelerate scientific and technological progress? If so, what are the most effective ways to discover the young Einsteins among us?

Knowledge on methods to identify individuals especially capable of making groundbreaking contributions to science and technology is especially valuable, considering that supporting talented youth is expensive and thus needs to be targeted. Currently, practitioners tend to use school grades, exam scores, and competition performance, but there is little systematic evidence on how well these indicators predict success in STEM careers.

One place to start collecting systematic evidence on how early signals correlate with participation and excellence in STEM are the International Science Olympiads. These international competitions bring bright high school students worldwide to tackle complex problems in fields such as mathematics, physics, chemistry, and informatics. More than just a test of knowledge, these contests aim to measure the participants' creative thinking and problem-solving abilities.

Success stories of Olympiad alumni abound. In the realm of technology, the creator of the Python language is an IMO alumni and half of OpenAI's founding team had participated in science Olympiads.[1] In pure mathematics, Grigori Perelman's, a gold medalist and prefect scorer at the IMO in 1982, went on to resolve the Poincaré conjecture, a complex mathematical problem unsolved for over a century.

In this paper, we study the International Science Olympiads and the extent to which their results are predictive of participation and excellence in STEM. From a research perspective, a major hurdle is the lack of relevant data connecting early academic achievements to later career outcomes, par-

[1] Guido van Rossum, an IMO bronze medalist, invented and developed the Python programming language. Among the founders of OpenAI, Greg Brockman was a silver medalist in the International Chemistry Olympiad, Wojciech Zaremba clinched silver at the IMO, and John Schulman reached finalist status for the US team in the International Physics Olympiad.

ticularly beyond undergraduate education. Additionally, labor force surveys offer minimal insight into early achievements. Further complicating the analysis is the fact that both youth excellence and STEM career success are inherently rare.

We leverage the fact that the results of international competitions are publicly announced to overcome the measurement challenges associated with studying the link between early signals and participation and excellence in STEM. By combining competition results with other types of publicly available data, including repositories of PhD theses and lists of scientific award winners, we are able to observe whether Olympiad participants become professional scientists and mathematicians, and whether they excel in these professions. We focus on the first five Olympiads to be established – in mathematics, physics, chemistry, informatics, and biology (which we collectively refer to as 'The International Science Olympiads'). We present three findings:

1. Among the top 500 Science Olympiad scorers annually, over 50% achieve PhDs, far exceeding the US average of 1% and the global average of 0.2%.

2. Among those who have won major scientific awards, 5% have taken part in Science Olympiads, rising to 13% for mathematics awards.

3. International Mathematical Olympiad (IMO) participants excel in both math and other fields – securing 80% of the scientific awards that are awarded to Olympiad alumni over time. Moreover, IMO medalists become scientific prize winners at thirty times the rate of undergraduate alumni from the world's top ten universities (seventy-five times for IMO gold medalists).

Overall, our results suggest that high school competitions, in the form of the International Science Olympiads, may be an excellent way to discover promising STEM talent at an early stage. Moreover, results from mathematics competitions appear to be particularly good indicators of potential to excel in STEM. A direct implication of our findings is that Olympiad results could be used more extensively for purposes of allocating scholarships and other forms of early support. While Olympiads appear to be a useful tool for STEM talent discovery, we highlight several avenues for additional inquiry. These include (1) exploring why mathematics competitions appear to be more predictive of STEM excellence, (2) developing a better understanding of whether (and, if so, why) high school competition excellence might have a causal effect on scientific achievements, (3) gath-

ering more systematic evidence on how other signals of early ability (such as grades, entrance exam scores, etc.) forecast an individual's likelihood of later excellence in STEM fields.

This paper relates to the discussion of policy options to support science and innovation (Freeman & Van Reenen 2009; Stepan 2010; Bloom, Van Reenen & William, 2019), and especially those related to the supply of scientists (Freeman 1975; Shu 2015, Toivanen & Väänänen 2016, Bell et al. 2019a, 2019b, Agarwal & Gaule 2020).[2] This paper is particularly relevant to one policy tool – financial support for aspirant scientists. While some research has evaluated the effect of such support at the doctoral or postdoctoral stage (Freeman 2005, Kahn & MacGarvie 2011, Jacob & Lefgren 2011, Graddy-Reed, Lanahan & D'Agostino 2021) little attention has been given on how to target the scholarships. Our paper suggests one effective way to target financial support to bright young individuals who are especially capable of advancing the knowledge frontier.

Closest to this paper are Aghion et al. (2017), Agarwal & Gaule (2020) and Agarwal et al. (2023). Among other results on parental income, Aghion et al. (2017) document that IQ in early adulthood (measured by an army entrance test) is correlated with the probability of inventing: being in the top decile of the IQ distribution increases the probability of inventing by 2-3 percentage points (compared to a sample mean of 1 percentage point). While this is not the focus of the paper, one possible reading of their results is that IQ tests could be used to identify individuals with a higher potential to invent. Agarwal & Gaule (2020) follow the careers of IMO medalists and document that those from developing countries are systematically less likely to engage in knowledge production. An important side result is that IMO gold medalists are fifty times more likely to win the Fields medal than PhD graduates from top 10 universities. While this result suggests – consistent with the present paper – that IMO medalists are especially capable of advancing the knowledge frontier, it is based on a single prize, the Fields medal, and involves no comparison with other Olympiads.[3] Finally, Agarwal et al. (2023) document that IMO medalists are more productive when they emigrate, and in particular, when they emigrate to the U.S. They also provide evidence that finance constraints prevent a substantial fraction of them to study in their preferred destination

[2] Whereas the literature has tended to emphasize the role of pull incentives in the form of salaries (Freeman 1975, Shu 2015) or taxation (Bell et al. 2019a), a number of recent studies take a broader view of factors shaping the supply of scientists ranging from the supply of education (Toivanen & Väänänen 2016), exposure to innovation (Bell et al. 2019a, 2019b) and beliefs (Ganguli, Gaule & Vuletic 2022).

[3] One particular issue with the Fields medal is that, since its eligibility is restricted to individuals 40 years of age or younger, it may favor precocious individuals – just as the high school competitions would.

– the U.S. or the U.K. The existence of financial constraints provides a rationale for scholarships, but in turn opens the question of targeting support.

The rest of this paper is organized as follows. In section 2, we provide brief background information on the International Science Olympiads. We describe our data collection strategy data in section 3 before presenting the results in section 4. Section 5 discusses and concludes.

2 The International Science Olympiads

The International Science Olympiads are competitive events designed to challenge and recognize high school students' skills in various scientific disciplines, including mathematics and the natural sciences. After a grueling multi-stage selection process, students who have excelled in regional and national competitions qualify to participate at the international level.[4] The international competitions themselves involve teams of 4 to 6 students who travel to a central location and compete to solve a selection of specially curated problems. While the competitions involve a team dimension with team rankings, students compete individually and earn individual recognition.

In this paper, we focus on the Olympiad in mathematics, physics, chemistry, informatics, and biology which were the first five Olympiads to be established. The International Mathematical Olympiad was established in 1959 and was followed by the International Physics Olympiad (1967), the International Chemistry Olympiad (1968), The International Olympiad in Informatics (1989) and the International Biology Olympiad (1990). While the International Olympiads exhibit some differences between each other in terms of team size, examination format, or number of participation allowed, they otherwise share many features. We provide additional information on each Olympiad in the appendix.

(Insert Figure 1 about here)

Figure 1 displays the number of participants in each Olympiad over time, as well as the number of countries participating in at least one Olympiad over time. From modest beginnings, the number

[4] In the U.S., for instance, students scoring in the top centiles of the (free to enter) American Mathematics Competitions are invited to the American Invitational Mathematics Examination, the top 500 scorers of which qualify for the USA Mathematical Olympiad. Tests results and training camps are then used to determine the team of six students that the US will send to the International Mathematical Olympiad.

of participants across all Olympiads grew slowly only the 1960s, 1970s and the first half of the 1980s. In 1985, there were only around 345 participants and 38 countries participating in at least one Olympiad. However, there has been remarkable growth in the 1990s, with the number of participants reaching 1,000 in 1996. Growth continued at a slower pace after, with 1800 participants across all five Olympiads in 2022. The growth in the number of Olympiad participants reflects a number of factors, including the creation of new Olympiads (biology and informatics in the early 90s) and more countries participating.

The reach of the national competitions feeding international ones varies across countries and subjects. At one extreme, 18 million Brazilian students, or 90% of school-age students, participate in OBMEP, the national Olympiad for public schools. In other countries and subjects, participation in the Olympiads may be down to a limited number of enthusiasts. However, most countries run competitive selection processes involving thousands of students.

3 Data

A key challenge in studying how early achievements-such as high school competition results-predict career outcomes is the absence of relevant data. Ideally, we would like to follow large cohorts of students as they progress through school and beyond, observe those who contribute to science and engineering, and relate these contributions to characteristics observable in high school, such as grades and competition results. Unfortunately, obtaining or constructing such data is challenging due to privacy and record linkages issues. We built an original dataset with the goal of providing a window into the careers of Olympiad medalists and the early achievements of scientific prize winners. Our data collection effort leverages publicly available information, including the fact that the International Science Olympiads publicly announce individual results. Using the website of Olympiads, we constructed a database of all past participants of the five Olympiad over the period between 1959 to 2022. Accounting for multiple participations, our database includes 37,244 individuals from 141 countries.

Scientific Prize Winners. To investigate the long-term achievements of these early science competition participants, we then created a list of major scientific prizes. Prizes are an attractive measure of contribution to science as they are awarded following careful selection procedures and

reflect the contributions of scientists as perceived by their community. We constructed a list of 126 awards (and their recipients) across mathematics, physics, chemistry, computer science, biology, and engineering.[5] The Nobel prize is naturally included in this list, along with the Wolf prizes, the Abel prize, and many others.

Our data includes 2,125 prize winners who have won one or more prizes from 2010 to 2023. We match those to their Olympiad participations records.[6] We also collect information on the university where the prize recipients have obtained their undergraduate degrees. This will enable us to compare how Olympiad medalists compare to undergraduates from top universities in their propensity to win scientific awards.

Doctoral Education. Besides long-term contribution to science, we were also interested in who becomes professional scientist, which we proxy through getting a PhD. Because collecting this type of data is costly, we focused on a subsample of 500 individuals, including the top 100 performers from each of the five major International Science Olympiads in 2000. We collected information on these individuals using multiple sources[7] and matched the profiles to Olympiad participants and. For comparison purposes, we also computed the percentage of individuals earning PhD degrees in the 1982 birth cohort in representative countries from different income groups.

4 Results

This section is divided into two parts. We first study the extent to which being a high scorer at one of these Olympiad is predictive of getting a PhD, as a proxy for becoming a professional scientists or mathematician. Second, we investigate which type of Olympiad backgrounds is predictive of making significant contributions to science.

[5] After careful consideration and cross-validation, we selected these awards based on their daily views on the corresponding Wikipedia pages. In doing so, we follow the approach of Ma & Uzzi (2023) in their systematic study of scientific prizes.

[6] While we need to take care of name disambiguation issues, matching the two datasets is relatively straightforward because we know both the population of International Science Olympiad participants and the population of prize-winners.

[7] The sources we use include but not limited to Wikipedia, Google Scholar, ORCID, Open Access Theses and Dissertations (OATD), ProQuest Dissertations & Theses Global (PQDT), ResearchGate, Linkedin, Facebook, university and institute websites, and local news.

4.1 Who Becomes a Professional Scientist or Mathematician?

We now present results on the extent to which top Olympiad scorers become professional scientists or mathematicians, as proxied by getting a doctoral degree. This exercise is based on following the top 100 scorers in each of the five major Olympiads, as described in the data section. As shown in Figure 2 panel A and B, 58 percent of Olympiad top scorers obtain degrees, rising to 64 percent in rich countries. For comparison purposes, only around 1 percent of the general population gets doctoral degrees in the US or in the OECD at large, and less than 0.2 percent worldwide. Olympiad top scorers are thus at least 50 times more likely to get a PhD than a member of the general population.

(Insert Figure 2 about here)

Being an Olympiad top scorer is an excellent predictor of getting a PhD not just in rich countries but also in poorer ones, where getting a PhD degree is an even rarer outcome. Around 59% of top Olympiad scorers in upper middle income (panel B) and 53% in low and lower middle-income countries (panel C) get a PhD degree. For comparison purposes only 0.28% of Brazilians and 0.08% of Indians get PhD degrees.

While this is not our main focus, we note that the share of Olympiad participants getting a PhD degree tends to be lower in poorer countries. In low- and lower middle-income countries, 53 percent of Olympiad top scorers get PhD compared to 59 percent in upper middle-income countries and 64 in high-income countries. This is consistent with the finding of Agarwal & Gaule (2020) that IMO participants from developing countries are less likely to become knowledge producers. However, our results here are noisier since we focus on a single cohort of participants.

We observe some differences across Olympiad types in the extent to which being a top scorer correlates with getting a doctoral degree. A higher share of top scorers in the mathematics, physics and chemistry Olympiad get doctoral degrees (65, 66 and 66 percent respectively), compared to biology and informatics (48 and 46 percent). We discuss possible reasons for these differences in the discussion section.

4.2 Who Advances the Knowledge Frontier

We are particularly interested in understanding the extent to which Olympiad prize winners contribute to advancing the knowledge frontier. To measure this empirically, we use a database of scientific prize-winners custom built for this project (see data and methods section). Overall, we find that that 5% (107 out of 2,124) of scientific prizes awarded in our sample have been won by participants from one of the five International Science Olympiads. Two pieces of contexts are useful to put this number in perspective. First, the popularity of the Olympiads is a relatively new phenomenon (as discussed in part 4. A.) and scientific achievements naturally lag high-school age results by quite a few years. Second, the number of individuals participating in the international Olympiads is in the hundreds (or low thousands in recent years) – we consider relative chances of winning an award later in this section.

(Insert Figure 3 about here)

Distinguishing between the competition individuals attended reveals that IMO participants won the lion's share (82%) of the awards won by Olympiad participants. That reflects the success of IMO participants in winning mathematics awards (12.6 percent of mathematics prizes have been won by IMO participants). However, IMO participants have also been successful in winning prizes outside mathematics – winning 11 prizes, more than the total number of prizes won by IOI or IPhO participants in any field (10 and 9 respectively). We could not find any award won by participants from the biology Olympiad and only 1 from the chemistry Olympiad participants. We complement the evidence on the number of awards won by considering the ratio of awards to the number of individuals having participated in a competition. We compute these numbers separately for all five competitions and metal category (gold, silver and bronze medal). For comparison purposes, we also consider the number of awards by undergraduate alumni from 10 leading global universities, as well as the number of undergraduate alumni form these universities.

(Insert Figure 4 about here)

Figure 4 displays the share of alumni winning scientific prizes for the alumni from leading global universities and International Science Olympiads alumni. Among the leading global universities,

the alumni from Caltech are the most successful with one in 862 going on to win a major scientific prize, followed by Harvard and MIT.

Alumni from certain International Science Olympiads win at a much higher rate. Among IMO gold medalists, for instance, 1 in 56 goes on to win a major scientific prize, or one in roughly 150 for IMO silver medalists or IOI bronze medalists. IMO medalists become scientific prize winners at thirty times the rate of undergraduate alumni from the world's top ten universities (seventy-five times for IMO gold medalists). Not all Olympiad medals are predictive of winning a major scientific prize: in biology and chemistry, we can hardly find any example of an Olympiad alumni winning a major scientific prize and correspondingly the share of alumni winning a prize is zero (except for chemistry gold).

(Insert Table 1 about here)

Table 1 analyzes the propensity of winning a scientific prize for Olympiad medallists and undergraduate alumni from top 10 universities in a regression format.[8] Medalists from one of the five international olympiads is associated are around 10 times more likely to win a scientific prize than an undergraduate alumni from 10 leading global universities (column 1). However, this masks substantial hetereogeneity across medal categories and Olympiads. Gold medallists in one of the Olympiads are 26 times more likely to win a major scientific prize than the undergraduate alumni; compared to 4.5 times for bronze medallists. Moreover, IMO and IOI medals are strongly associated with winning a prize whereas IChO and IBO Olympiads are not (column 3). When distinguishing both medals and Olympiad categories, an IMO gold is associated with a 70x higher chance of winning scientific prize, following by IMO Silver (24x) and IOI Gold (17x).

The results described above suggest that medalling in an Olympiad is not necessarily predictive of winning a subsequent scientific award. However, winning a gold medal and/or medaling in a mathematics competition is strongly predictive of winning a subsequent scientific award.

[8]These regressions are run on the union of the set of Olympiad medallists (n=21,243) and the set of undergrad alumni from 10 leading global universities (n=813,481). The dependent variable is an indicator for winning one or major scientific prizes. The variables of interest are indicator variables for different types of medal in one of the International Science Olympiads. The omitted category/reference is undergraduate alumni from 10 leading global universities. To facilitate the interpretation of coefficients, the dependent variable is rescaled by a factor 1/0.000241 so that its mean is 1 for the undergraduate alumni from 10 leading global universities.

5 Discussion

Surprisingly little evidence exists on whether and how it might be possible to identify at an early-stage individuals who are especially capable of advancing the knowledge frontier. In this paper, we have considered the case of competitions for high school competitions as a tool to discover such individuals. We have documented that excellence in the Olympiad is a very good predictor of getting a doctoral degree as well as receiving scientific awards for outstanding contributions to science and mathematics. While Olympiad results seem to be predictive of engaging in a scientific career across Olympiad types (mathematics, physics, informatics, biology and chemistry), the results for are more nuanced for scientific awards, with mathematics competitions being especially predictive and others (chemistry and biology) hardly so.

Science and technology policy typically focuses on individuals who are already engaged in scientific careers, in the form of grants, fellowships, prizes, or other support. However, there is increased recognition that intervening earlier to shape the pool of individuals who enter scientific careers may be desirable and important. Policymakers may be interested in increasing participation from historically underrepresented groups for its own sake. Alternatively, early interventions may also be motivated by efficiency considerations (for instance increasing the quality of the STEM workforce). In either case, targeting early interventions to the right individuals is critical, especially when such interventions are expensive on a per-individual basis. The results of our papers suggest a practical way (i.e. competitions) to identify individuals at an early stage for the purposes of allocating scarce university places, fellowships, various forms of exposure to the STEM careers, or other forms of support.

While competitions appear to be a useful tool to identify talented individuals with the potential to excel in STEM, they do suffer from several limitations. First, some talented individuals may dislike the competitive elements of Olympiads. This could be the case for women, for instance, who have been historically underrepresented in them.[9] Second, many talented youngsters around the world have limited opportunities to participate in the Olympiads, due to lack of awareness, cost considerations or sheer lack of competitions in their country or region. In Africa, for instance, the

[9]Historically less than 10 percent of participants to the IMO have been women. There may be ways to make competitions more inclusive and appealing to women for instance through female-only competition. Ultimately, though, Olympiads are competitions which may be unappealing to certain individuals and groups.

reach of competitions remains generally limited. In other developing countries, participation fees to participate in Olympiad may be a deterrent. Third, the time invested in preparing for competitions comes with opportunity costs in terms of other activities, including academic enrichment activities.

We conclude by noting another a couple of open questions. One may wonder why mathematics competitions are more predictive of the ability to push the knowledge frontier. A simple explanation is that the process of knowledge production in other disciplines may be more reliant on factors, for instance luck and physical capital. A challenge for that explanation is that we find youngsters who excel in mathematics competitions go on to win scientific prizes not just in mathematics but also in other domains. Another possibility is that the types of problems asked in mathematics competitions are better at capturing certain dimensions of intelligence and ability which in turn correlate with scientific excellence. A third, more prosaic, explanation, is that mathematics competitions tend to reach a larger number of kids, as they have been established earlier, have stronger traditions, and only require pen and paper to train and participate. As a consequence, mathematics Olympiad medalists are selected from a bigger pool of individuals (than medalists from other Olympiads) and may be more talented as a result. Further research to probe into these various explanations would be useful.

Another question is whether the link between high-school competitions results and scientific excellence is simply a correlation or whether it also reflects a causal effect. Acquiring advanced problem-solving skills as a teenager may be helpful in the production of the knowledge at later life stages. Alternatively, those who do well in competitions may be encouraged to continue in the scientific path, either through their own self-confidence or by receiving access to opportunities, such as scholarships to study in top universities.[10] In this paper, we have been interested in predictors of scientific excellence, and made the case that these are useful in themselves. However, if there is some causality between high school competitions results and subsequent scientific excellence, that would provide an additional rationale (besides talent identification) for policymakers to support the development of such competitions. Finally, much remains to be learned about the effectiveness of

[10]Based on a regression discontinuity design and data from the Brazilian public mathematics Olympiad, Moreira (2019) finds that winning a honourable mention increases the confidence not just of the winners but also of their peers at school. Also using a regression discontinuity design around IMO medal thresholds, Agarwal & Gaule (2020) do not find that getting a better medal at the IMO increases the chance of doing a PhD. There is even less evidence on the value of training for the IMO and similar competitions.

other forms of talent identification, both by themselves an in comparison with competitions. Besides collecting grades at school, many school students go through high-school exit exams or university entrance exams. To what extent do these results correlate with future excellence? Relatedly, could students' extracurricular activities (e.g. chess, music, summer projects) be used to detect students with very high potential? Better knowledge on identifying talent is key to ensuring that talent is properly identified and nurtured and reaping the associated benefits in terms of science, innovation and growth.

References

Agarwal, R., & Gaule, P. (2020). "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?" American Economic Review: Insights, 2(4): 409-424.

Agarwal, R., Ganguli, I., Gaule, P., & Smith, G. (2023). "Why US Immigration Matters for the Global Advancement of Science". Research Policy, 52(1): 104659.

Aghion, P., Akcigit, U., Hyytinen, A., & Toivanen, O. (2017). "The Social Origins of Inventors" NBER WP No 24110. National Bureau of Economic Research.

Bell, A. M., Chetty, R., Jaravel, X, Petkova, N., & Van Reenen, J. (2019a). "Do Tax Cuts Produce More Einsteins? The Impacts of Financial Incentives versus Exposure to Innovation on the Supply of Inventors." Journal of the European Economic Association, 17(3): 651-677

Bell, A. M., Chetty, R., Jaravel, X., Petkova, N., & Van Reenen, J. (2019b). "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." Quarterly Journal of Economics, 134(2): 647-713.

Bloom, N., Van Reenen, J., & Williams, H. (2019). "A Toolkit of Policies to Promote Innovation." Journal of Economic Perspectives, 33(3): 163-184.

Choudhury, P., Ganguli, I., & Gaule, P. (2023). "Top talent, Elite Colleges, and Migration: Evidence from the Indian Institutes of Technology." Journal of Development Economics, 156(2023): 103120.

Jacob, B. A., & Lefgren, L. (2011). "The impact of NIH postdoctoral training grants on scientific productivity." Research Policy, 40(6): 864-874.

Freeman, R. B. (1975). "Supply and salary adjustments to the changing science manpower market: Physics, 1948-1973." American Economic Review, 65(1): 27-39.

Freeman, R., & Van Reenen, J. (2009). "What if Congress Doubled R&D Spending on the Physical Sciences?" Innovation Policy and the Economy, 9(1): 1-38.

Graddy-Reed, A., Lanahan, L., & D'Agostino, J. (2021). "Training across the academy: The impact of R&D funding on graduate students." Research Policy, 50(5): 104224.

Freeman, R. B. (2005). "Fellowship stipend support and the supply of science and engineering students: NSF graduate research fellowships." American Economic Review, 95(2):61-65.

Ganguli, I., Gaule, P., & Čugalj, D. V. (2022). "Chasing the Academic Dream: Biased beliefs and Scientific Labor Markets. " Journal of Economic Behavior & Organization, 202: 17-33.

Kahn, S., & MacGarvie, M. (2011). "The Effects of the Foreign Fulbright Program on Knowledge Creation in Science and Engineering." In The Rate and Direction of Inventive Activity Revisited (pp. 161-197). University of Chicago Press.

Ma, Y., & Uzzi, B. (2018). "Scientific prize network predicts who pushes the boundaries of science." Proceedings of the National Academy of Sciences, 115(50), 12608-12615.

Moreira, D. (2019). "Success Spills Overs: How Awards Affect Winners and Peer's Performance in Brazil". Mimeo, UC Davis.

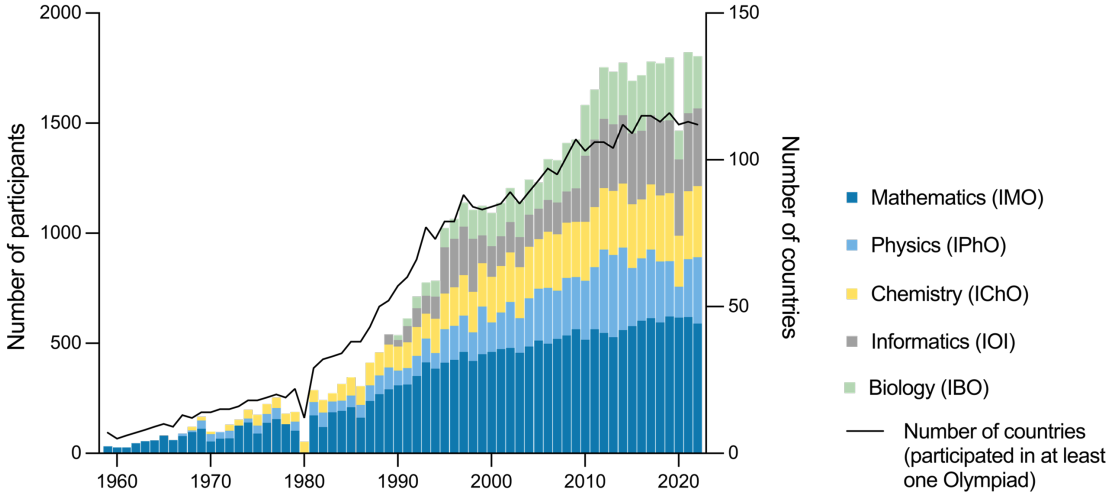
Shu, P. 2015. "Are the 'Best and Brightest' Going into Finance? Career Choice and Skill Development of MIT Graduates." Harvard Business School Working Paper #16-067.

Stephan, P. E. 2010. "The Economics of Science." In Bronwyn H. Hall and Nathan Rosenberg (Eds.), Handbook of The Economics of Innovation, pp. 217-273. Amsterdam: North-Holland.

Toivanen, O.; & Väänänen, L. 2016. "Education and Invention." Review of Economics and Statistics, 98(2): 382-396.

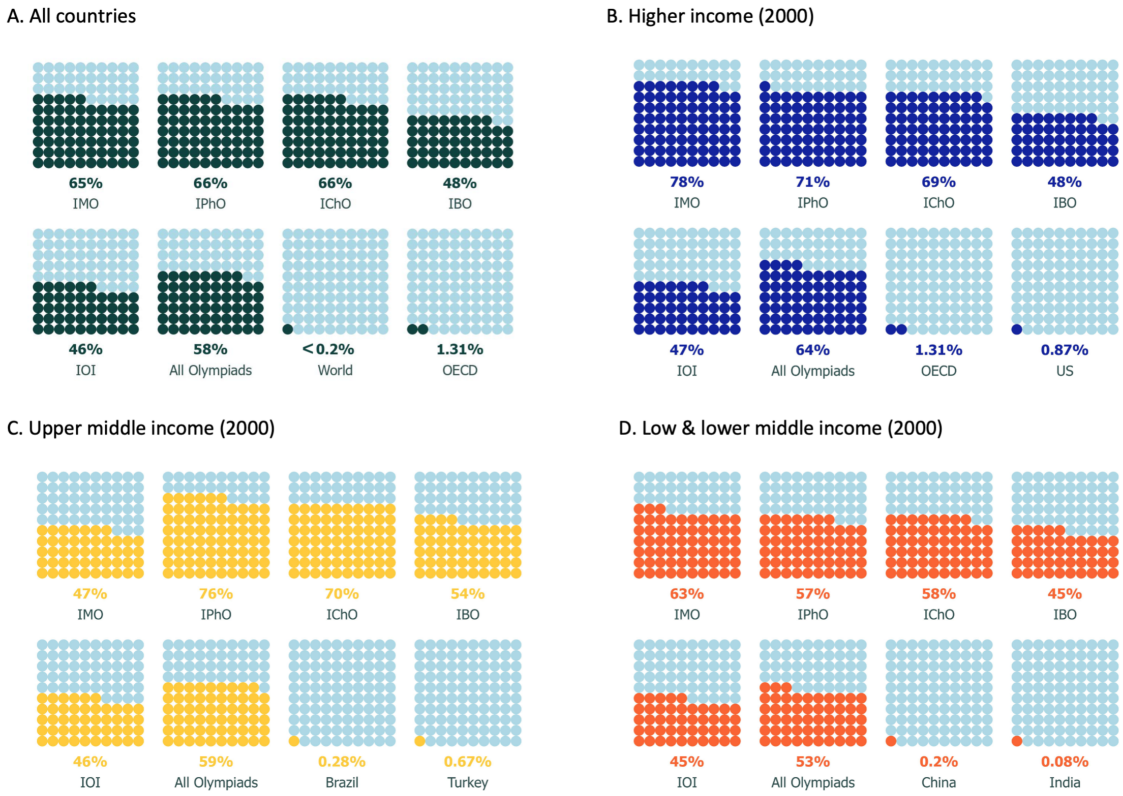
Tables and Figures

Figure 1: Participation in the International Science Olympiads



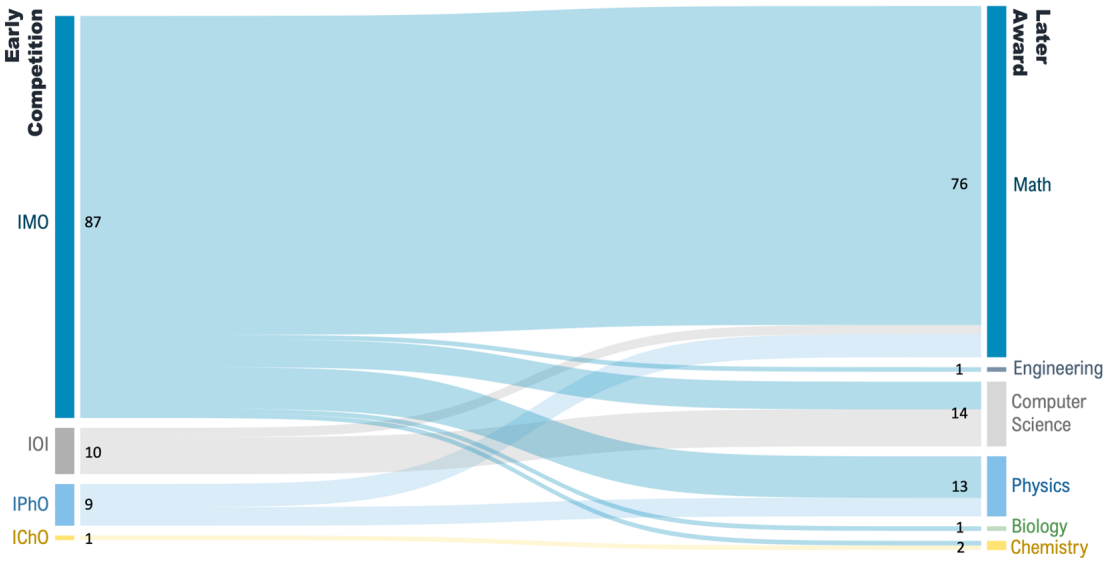
Notes: This Figure includes data for all participants and participating countries of the five major International Science Olympiads (IMO, IPhO, IChO, IOI, and IBO) from 1959 to 2022. The stacked bars demonstrate the number of participants in each Olympiad and the line shows the number of countries participated in at least one Olympiad.

Figure 2: Obtaining a Doctoral Degree



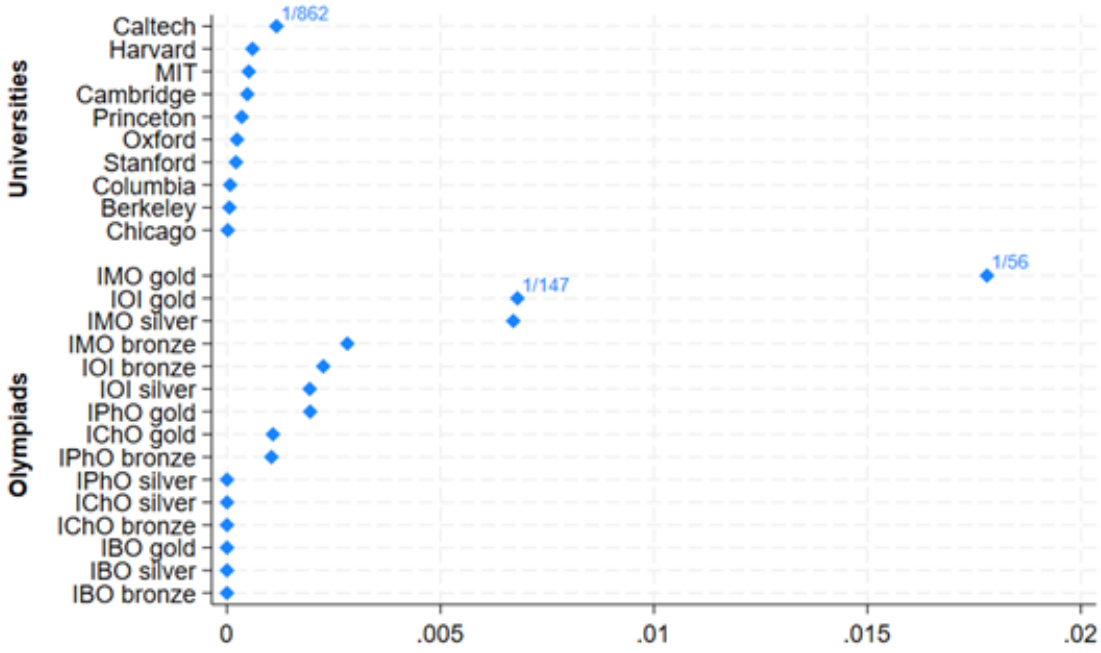
Notes: This Figure tracks the education of the top 100 scorers in each of five major Olympiads (held in 2000), with the general population as a comparison groups.

Figure 3: Scientific Prizes won by International Science Olympiads Alumni



Notes: The Figure shows the number of scientific prizes won between 2010 and 2023 by alumni from the major International Science Olympiads. The width a given band represent the number of prizes won by alumni from a given olympiad in a given discipline. The numbers in parentheses on the left represent the total numbers of prizes (across all disciplines) won by alumni from each olympiad. The numbers in parentheses on the right represent the total number of prizes won in each discipline by alumni from any of the five Olympiads. The International Biology Olympiad (IBO) is not shown on the graph as no scientific prizes have been won IBO alumni in our data.

Figure 4: Share of Alumni Winning Scientific Prizes



Notes: This Figure was constructed by (1) computing the number of scientific prize winners (awarded between 2010 and 2023 that are alumni from 10 leading global universities or one of the major international science Olympiads (distinguishing by olympiad and medal category), (2) computing the number of undergraduate students (enrolling between 1980 and 2009) and the number of Olympiad alumni (participating between 1980 and 2003) within each olympiad and medal category; and (3) taking the ratio between the numbers from (1) and (2).

Table 1: Propensity to Win a Scientific Prize

	(1) Overall		(2) Medal Types		(3) Olympiads		(4) Both	
Any medal	10.33***	(0.50)						
Gold Medal			26.64***	(1.10)				
Silver Medal			8.39***	(0.86)				
Bronze Medal			4.50***	(0.73)				
IMO medal					29.94***	(0.92)		
IOI medal					10.25***	(1.33)		
IPhO medal					1.90*	(1.11)		
IChO medal					-0.10	(1.10)		
IBO medal					0.01	(1.15)		
IMO Gold							69.94***	(2.13)
IMO Silver							23.95***	(1.52)
IMO Bronze							9.06***	(1.23)
IOI Gold							17.16***	(3.07)
IOI Silver							4.29**	(2.18)
IOI Bronze							6.16***	(1.78)
IPhO Gold							3.00	(2.27)
IPhO Silver							-1.63	(1.91)
IPhO Bronze							2.84*	(1.56)
IChO Gold							3.71	(2.39)
IChO Silver							-1.16	(1.79)
IChO Bronze							-1.03	(1.47)
IBO Gold							-0.91	(3.02)
IBO Silver							-0.89	(2.19)
IBO Bronze							-0.95	(1.78)
Obs.	834,724		834,724		834,724		834,724	
R2	0.0005		0.0009		0.0013		0.0023	
Mean of DV (raw)	0.000241		0.000241		0.000241		0.000241	
Mean of DV (adj)	1		1		1		1	

Notes: These regressions are run on the union of the set of Olympiad medallists (n=21,243) and the set of undergrad alumni from 10 leading global universities (n=813,481). The dependent variable is an indicator for winning one or major scientific prizes. The variables of interest are indicator variables for different types of medal in one of the International Science Olympiads. The omitted category/reference is undergraduate alumni from 10 leading global universities. To facilitate the interpretation of coefficients, the dependent variable is rescaled by a factor 1/0.000241 so that its mean is 1 for the undergraduate alumni from 10 leading global universities. Standard errors in parentheses * p < 0.1, ** p < 0.05, *** p < 0.01

A Further Background on the International Science Olympiads

The International Science Olympiads are a series of prestigious international competitions held annually for high school students. They cover a wide range of scientific disciplines and aim to challenge and recognize exceptional young minds in various scientific fields. These Olympiads typically include formal sciences (e.g., mathematics and computer science), natural sciences (e.g., physics, chemistry, and biology), and sometimes social sciences as well. During the Science Olympiads, teams of students representing different countries compete against each other through rigorous exams and practical problem-solving tasks. The competitions are organized at the national level first, and the top-performing students from each country then advance to the international stage. In this paper, we mainly focus on five well recognized International Science Olympiads in mathematics, physics, chemistry, computer science, and biology.

A.1 IMO

The International Mathematical Olympiad (IMO) is the World Championship Mathematics Competition for High School students and has been held annually in a different country since 1959. It started in East European countries and has gradually expanded to over 100 countries. A participating country may send a team of up to six students under the age of 20 and not enrolled in a tertiary institution. In the competition, there are six problems worth seven points each for a maximum total score of 42 points. The problems cover a variety of secondary school mathematics topics, such as geometry, number theory, algebra, and combinatorics, with no prior knowledge of higher mathematics required. Competitions are held over two consecutive days, during which contestants must solve three problems in four-and-a-half hours. Prizes are awarded to approximately half of the top-scoring contestants, with gold, silver, and bronze medals distributed roughly 1:2:3.

A.2 IPhO

Inspired by the IMO, the first International Physics Olympiad (IPhO), which is the World Championship Physics Competition for High School students, was held in Poland in 1967. It has gradually expanded to over 80 countries. Each national delegation is made up of at most secondary school

students. The main aim of IPhO is to test the highest level of knowledge, critical thinking, problem solving, right practices of presentation and analysis, and hands-on skills in theoretical and experimental physics. The competition consists of two examinations, one theoretical examination lasts 5 hours and consists of three questions and one practical examination may consist of one or two experimental tasks which together last five hours. Gold, silver, and bronze medals are rewarded to the top 8%, 25%, and 50% of participants and an honourable mention or better is awarded to the top 67%.

A.3 IChO

The International Chemistry Olympiad (IChO) is a prestigious international chemistry competition for high school students. The first IChO was held in 1968, and it has gradually expanded to over 80 countries. Every year, each national team competes with up to 4 students. In the competition, a five-hour laboratory practical exam and a five-hour written theoretical examination are held on separate days to assess participants' knowledge and skills in chemistry. A gold medal is awarded to the top 12% of students, a silver medal to the next 22%, a bronze medal to the next 32%, and an honourable mention is awarded to the top 10% of participants who did not receive a medal.

A.4 IOI

The International Olympiad in Informatics (IOI), an annual competitive programming competition, is one of the International Science Olympiads for secondary school students. The first IOI was held in 1989 in Bulgaria to stimulate interest in informatics (computing science) and information technology. It now has more than 80 countries participating every year. Each participating country can send up to four students to take part in the contest. Students are typically given three problems to solve in C++ in five hours on each competition day. All problems from both competition days are summed up separately for each individual, and medals are awarded based on their relative total scores. The top half of the participants are awarded medals, with a ratio of approximately 1:2:3 for gold, silver, and bronze.

A.5 IBO

The International Biology Olympiad (IBO) is a biological competition for pre-university students under the age 20, and is one of the most well-known International Science Olympiads. The first IBO was held in Czechoslovakia in 1990 and it have gradually expanded to include more than 70 participating countries. Every year, participating countries send the four winners of their National Biology Olympiad to the IBO and the participants cannot compete more than two times in the competition. During IBO, students have to participate in Theoretical Exams (50%) and Practical Exams (50%). The Theoretical Exams include approximately 100 questions, which cover topics including cell biology, plant anatomy and physiology, animal anatomy and physiology, ethology, genetics and evolution, ecology, and biosystematics. The practicals usually consist of three to four laboratories, with content domains announced approximately six to twelve months before the IBO. By conducting experiments and analysing data in the lab, students practice biological skills. The competitors are ranked based on their equally weighted scores for the theoretical and practical tasks according to the t-score of individual theoretical and practical tasks. The top 60% of participants are awarded medals, with a ratio of approximately 1:2:3 for gold, silver, and bronze, and the top 25% of non-medallists are awarded a certificate of merit.

B Constructing The Doctoral Education Database

We focused on a subsample of 500 individuals, comprising the top 100 participants from the five Olympiads in 2000. To link their profiles to competition participants and chess players, we used various sources and applied the following matching criteria:

1. We checked if the individual's CV or profile mentioned their participation in Olympiads.
2. For those who began university studies between 2000 and 2003 or obtained a PhD between 2008 and 2015, majoring in STEM or Economics (Finance), and had a relatively unique name (not yielding over 1,000 results on Linkedin), we considered a potential match.
3. We ensured that the profile photo corresponded to the official photo taken during the competitions.

Profiles were matched to Olympiad participants if they satisfied any of the above criteria (i, ii, or iii). As a proxy for medium-term outcomes in science, we utilized the attainment of a doctoral degree, which includes degrees like Doctor of Philosophy (PhD), Juris Doctor (JD), and Doctor of Public Health (DrPH). Individuals without a matched online profile or lacking educational information on their profile were coded as having no doctorate degree

Countries or regions in our sample were categorized into four income groups—low-, lower-middle-, upper-middle-, and high-income groups—based on their GNI per capita in 2000, calculated using the World Bank Atlas methods. Due to the limited number of Olympiad participants and chess players in the low- and lower-middle-income groups, we combined these two groups into one, resulting in three income groups for comparison. We considered the US and 26 OECD countries as representatives of the high-income group, Brazil and Turkey for the upper-middle-income group, and China and India for the low- and lower-middle-income group.

In our analysis, we made the assumption that individuals typically obtain their doctoral degree around age 28. We then calculated the probability of obtaining a doctoral degree among those born in 1982 for representative countries from different income groups. To calculate the chance of obtaining a doctoral degree, we divided the number of people who obtained a doctoral degree in 2010 by the total number of births in 1982 for each respective country. The results of these calculations are presented in Table S1, along with the relevant sources for each country's data. In the US, a large share of PhD graduates are international students. In the US, a large share of PhD graduates are international students. Thus, for the US, we use the number of US citizens obtained a doctoral degree instead of the number of doctoral graduates in the country. For India, we did not find the number of doctoral graduates for the year 2010, so we used the data for the academic year 2011-12 instead.

B.1 Constructing the Prize Recipients Database

Apart from studying who gets a PhD degree, we were also interested in the extent to which Olympiad medalists make groundbreaking achievements in science. To measure groundbreaking achievement, we compiled a list of scientific prizes.

List of prizes. The selection of these prizes was based on their daily views on the corresponding

Wikipedia page, using Pageviews Analysis as a reference. We included an award in the list if its corresponding Wikipedia page had an average of no less than 5 daily views (2 for early or mid-career awards) from 1st July 2015 to 31st October 2022. Additionally, these awards needed to have more than 10 recipients since their establishment and should be international, intercontinental, or eligible for individuals working in the UK or US. Due to the limited history of IOI and IBO, we restricted the awards to early and mid-career awards in Computer Science and Biology. Our analysis has resulted in a list of 124 later scientific awards selected from seven fields, including Mathematics, Physics, Chemistry, Computer Science, Biology, Engineering, and Economics, based on these criteria.

List of prize winners. After selecting the awards, we collected information about the award winners from 2010 to 2023. The list of later award winners was obtained from the official websites of the prizes and awards in November 2022. Only prizes or awards awarded in or before November 2022 were included in our sample, and any prizes or awards awarded after that date were not considered in the analysis. Our data includes 2,124 awards made to 1,744 distinct individuals from 2010 to 2023.

Data collection for prize winners. For each prize winner, we manually collected data on educational history, including undergraduate degree institution and year. We also coded whether they participated in the International Science Olympiads and with what results.

Alumni from leading global universities. To establish a group to compare Olympiad medallists with, we first defined a list of 10 leading global universities.[11] We then used public source to collect information on the number of undergraduate students enrolling in each of these universities across all majors.

[11] Based on the 2022 Academic Ranking of World Universities (ARWU), we selected the following top 10 universities are: Harvard University, Stanford University, University of Cambridge, University of California, Berkeley, Massachusetts Institute of Technology (MIT), Princeton University, University of Oxford, Columbia University, California Institute of Technology (Caltech), and University of Chicago. See http://www.shanghairanking.com/rankings/arwu/2022 for more information.