ONLINE APPENDIX

This appendix provides further information on Winner-Nominee comparability and reports the results from additional analyses that test the robustness of the results provided in the main manuscript.

A.1 Alternative model specifications: fixed effects

To further test the robustness of the results in Table 2 in the main text, we alternate between scientist-, article and field of study × year fixed effects, while keeping year fixed effects and the control variables *Article age* and *Career age* in all specifications, to mitigate a general upward time trend in citations. The results in Table A.1 indicate that the key coefficient (*Scientist collaborated with winner* × *Post award*) is largely stable across the alternative sets of fixed effects.

[Insert Table A.1 about here]

A.2 Alternative model specifications: Count model

Our main specification does not use a count model for practical reasons: overdispersion in the dataset and a very large number of fixed effects. In contrast, an ordinary least squares regression does not suffer from the issue of overdispersion, thereby alleviating the need to adjust standard errors for overdispersion. Furthermore, in contrast to count models, there are no computational constraints from adding a large number of fixed effects to an OLS regression. Finally, the interpretation of the economic effect in a classic difference-in-differences regression is straightforward. Still, given the count nature of our dependent variable, it is important to know whether our results hold in a count model. Table A.2 reports the results from a Poisson model that uses the number of citations to an article in a given year as

the dependent variable. We find positive and significant coefficients in both models on the interaction term. Moreover, the economic magnitudes are similar to those of the difference-in-differences linear regression model in the main analysis: Scientists who collaborated with Nobel Prize winners increase their citations in the post-award period by (exp(0.611)-1=) 84 percent in Model 1 and (exp(0.473)-1=) 61 percent in Model 2, relative to citations of the coauthors of nominees in the post-award period. This compares to increases of 59 percent and 29 percent respectively in the linear regression difference-in-differences model (see footnotes 10 and 11 in the manuscript).

[Insert Table A.2 about here]

A.3 Alternative control groups

Excluding coauthors that worked with winners and nominees. In our analysis, we assign a coauthor as part of the winner group if any of his/her collaborators ever won a Nobel Prize in Physics. 55 coauthors (or about 2.5 percent of coauthors) in our dataset worked with multiple nominees where one of the nominees also won the Nobel Prize. Consequently, in those 55 cases, a coauthor is assigned to the group of "winner coauthors" (i.e., the treatment group) even though another star among his/her collaborators was, at either an earlier or later date, also unsuccessfully nominated for the Nobel Prize. We address whether their inclusion influenced our results by repeating our analysis while excluding these 55 coauthors. As reported in Table A.3, the results remain virtually unchanged.

[Insert Table A.3 about here]

Propensity-score matched coauthors of nominees. We examine whether a change to the control group (i.e., to the articles by nominee coauthors) influences our results. Specifically, in Table A.4, models 1 and 2, we use nearest-neighbor propensity-score matching to

obtain articles by coauthors of nominees that are most similar to those of coauthors of winners. For each winner coauthor article, we identify a control article that is, in the event year, closest in article age and in the scientist's career age. To measure the post event year change, each control article selected in this manner is then retained in the datasets with all its pre- and post-years. We then repeat the main analysis from the manuscript with this modified control group. In both specifications, the coefficients of interest on *Scientist collaborated with winner* $\times$ *Post award* remain highly significant and similar in magnitude (0.017*** and 0.020***) to those reported in Table 4 (0.016*** and 0.013**). We conclude that modifying the control group of articles written by coauthors of nominees to better match the articles written by the coauthors of winners does not affect the findings reported in the manuscript's Table 2.

[Insert Table A.4 about here]

Matching nominees' time under consideration. An alternative approach to modifying the control group is to keep only the Nobel Prize nominees who are most comparable with the winners of the Nobel Prize in the event year (before then identifying their coauthors with their respective articles and citations). Hence, we examine whether accounting for the duration of nomination prior to winning – i.e., the nominee and the winner are both under consideration for the prize for a comparable amount of time – influences the effects of coauthoring with a winner. To do so, we compute for each nominee (in a given event year) the time since the first nomination. We then choose only the nominee who is closest in duration to the eventual winner in that event year.[1] Once we identify a nominee as the closest match in terms of time under consideration, we include all of his/her coauthors' articles and citations throughout the event period (1850-1975). We note that because a winner typically has several nominations before winning the Nobel Prize (on average, a winner in our dataset has five nominations), the procedure of keeping only the

[1]In the few cases when there is more than one winner, we take the average duration for the winners.

nominee who has the most similar time under consideration removes those nominees who are infrequently nominated. Models 3 and 4 in Table A.4 report the results. We find highly significant coefficients of interest that are similar in size (0.022*** and 0.014**) to those reported in Table 4 (0.016*** and 0.013**). In consequence, further tightening the match between nominees and winners does not affect the findings reported in the main text.

Restricting control group to "best" nominees. An alternative to selecting the nominee who is closest in time under consideration to the winner is to select the "best" (never-winning) nominee in a given event year. We use four different criteria for best nominee that leverage the unique context of our setting, in which we have information about each instance that a nominee was nominated and – in the Nobel Committee's unembargoed data – the identity of their nominators. Our "best nominee" criteria are:

1. the greatest number of times that a nominee had been nominated till a given event year;
2. the greatest number of distinct nominators till a given event year;
3. the greatest sum of authority scores of the distinct nominators till a given event year;
4. the greatest number of distinct nominators that were Nobel Laureates.

Using the first criteria ensures that only those nominees who are repeatedly under consideration for the Nobel Prize are kept in the model. This minimizes any quality-differences that may exist between winners, who are often nominated several times before winning, and those nominees (now removed) who are infrequently nominated. The second, third, and fourth criteria leverage the information that we have on the experts who nominate researchers for the Nobel Prize. For the second criterion, having more nominators should correlate with higher levels of recognition and may indicate that a nominee produced research that was almost as impactful as the research by a winner. The third criterion similarly taps into nominator

recognition albeit with a different measurement, the authority score, which refers to the Nobel Committee's systematic ranking of nominator expertise.[2] The fourth criterion, whether a nominator was a Nobel Laureate, captures differences in nominator influence as the opinions of former laureates were given the most weight by the selection committee (Gallotti and De Domenico, 2019); thus, those nominated by former laureates had the greatest chance of winning in a given year. For all criteria, multiple nominees are kept if there is a tie between nominees.[3]

After determining the best nominee in an event year, we identify the coauthors of the best nominees and the winners, their articles, and the citations to those articles across the 1850-1975 event period. Table A.5 reports the results. Across all eight model specifications, the coefficient of interest on *Scientist collaborated with winner* × *Post award* remains positive, highly significant and – with values between 0.010*** and 0.025*** – similar in magnitude to those in the manuscript's Table 2. This provides further evidence that the models are robust to changes in the control group, and that the articles of the coauthors of winners receive more citations than those of the coauthors of nominees who came close to receiving but never actually won the Nobel Prize in Physics.

[Insert Table A.5 about here]

[2]The ranking, while implicit, is referred to as "nominator authority" in Crawford's census of Nobel Prizes in physics and chemistry (see Crawford (2002) and `https://www.nobelprize.org/nomination/physics/` for details).

[3]For instance, if two nominees had each been nominated 5 times until a given event year and all other nominees had been nominated fewer times, then both nominees were kept as both fit the criteria of best nominee.

References

Correia, S., P. Guimarães, and T. Zylkin. 2020. Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal* 20:95–115.

Crawford, E. 2002. *The Nobel Population 1901-1950: A Census of the Nominators and Nominees for the Prizes in Physics and Chemistry.* Tokyo, Japan: Universal Academy Press.

Gallotti, R., and M. De Domenico. 2019. Effects of homophily and academic reputation in the nomination and selection of Nobel laureates. *Scientific Reports* 9:17304.

Table A.1: Effect of collaboration on citations – Alternative fixed effect specifications

Dependent variable	*Log cumulative citations*							
Sample	Citations to articles published prior to award				Citations to articles published prior to award and first collaboration			
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post award	0.013***	0.012***	0.012***	0.011***	0.012***	0.012***	0.012***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Scientist collaborated with winner × *Post award*	0.020***	0.017***	0.017***	0.016***	0.016***	0.014***	0.014***	0.013**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	2,090,683	2,090,392	2,090,392	2,073,031	1,065,501	1,065,499	1,065,499	1,051,877
Number of clusters	41,433	41,142	41,142	40,746	17,764	17,762	17,762	17,507
R-squared	0.153	0.796	0.796	0.811	0.184	0.773	0.773	0.790
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	No	Yes	Yes	Yes	No	Yes	Yes
Article FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Field FE	No	No	Yes	n/a	No	No	Yes	n/a
Year FE	Yes	Yes	Yes	n/a	Yes	Yes	Yes	n/a
Field of study × Year FE	No	No	No	Yes	No	No	No	Yes

Notes: The table shows the results to alternative model specifications to those in Table 3 of the manuscript. Control variables *Article age* and *Career age*. Robust standard errors are clustered at the article level and reported in parentheses. In models without scientist fixed effect, the variable *Scientist collaborated with winner* is included but the coefficient for brevity unreported. $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.10$.

Table A.2: Effect of collaborations on citations – count model

Dependent variable	*Citations*	
Sample	Citations to articles published prior to Nobel Prize award	Citations to articles published prior to Nobel Prize award and first collaboration
Model	(1)	(2)
Post award	-0.555***	-0.805***
	(0.088)	(0.141)
Scientist collaborated with winner	0.611***	0.473***
× *Post award*	(0.178)	(0.158)
Observations	2,090,685	1,065,501
Number of clusters	41,435	17,764
R-squared	0.648	0.611
Control variables	Yes	Yes
Scientist FE	Yes	Yes
Article FE	Yes	Yes
Field of Study × Year FE	Yes	Yes

Notes: The table results the results for two Poisson count models. Observations are at the article-author-year level. The dependent variable, Citations, is the number of citations an article received in a given year. Robust standard errors are clustered at article level (41,435 clusters in Model 1; 17,764 clusters in Model 2) and reported in parentheses. $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.10$.

Table A.3: Effect of collaboration on citations — Alternative control groups

– Coauthors who worked with both winners and nominees are excluded –

Dependent variable	*Log cumulative citations*	
Sample	Citations to articles published prior to Nobel Prize award	Citations to articles published prior to Nobel Prize award & first collaboration
Model	(1)	(2)
Post award	0.011***	0.011***
	(0.002)	(0.002)
Scientist collaborated with winner × *Post award*	0.017***	0.014***
	(0.004)	(0.005)
Observations	2,047,159	1,034,252
Number of clusters	40,207	17,211
R-squared	0.812	0.791
Control variables	Yes	Yes
Scientist FE	Yes	Yes
Article FE	Yes	Yes
Field of study × Year FE	Yes	Yes

Notes: This table reports coefficients for two ordinary least square (OLS) regressions. Observations are at the article-author-year level. The dependent variable is the natural logarithm of one plus the number of citations that an article accumulated up to a given year (excluding any self-citations). Control variables include *Article age* and *Career age*. Robust standard errors are clustered at the article-level and reported in parentheses. $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.10$.

Table A.4: Effect of collaboration on citations — Alternative control groups

– Matching coauthor articles and nominee time under duration –

Dependent variable	*Log cumulative citations*			
Sub-sample	*Nominee coauthors propensity-score matched to winner coauthors*		*Coauthors of nominees with closest duration to winner in event year*	
Sample	Citations to articles published prior to Nobel Prize award	Citations to articles published prior to Nobel Prize award & first collaboration	Citations to articles published prior to Nobel Prize award	Citations to articles published prior to Nobel Prize award & first collaboration
Model	(1)	(2)	(3)	(4)
Post award	0.005*	-0.000	-0.001	0.001
	(0.003)	(0.004)	(0.002)	(0.004)
Scientist collaborated with winner × *Post award*	0.017***	0.013**	0.022***	0.015**
	(0.004)	(0.005)	(0.004)	(0.006)
Observations	918,582	464,087	1,180,868	579,160
Number of clusters	17,527	7,681	21,543	9,241
R-squared	0.798	0.774	0.804	0.782
Control variables	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes
Article FE	Yes	Yes	Yes	Yes
Field of study × Year FE	Yes	Yes	Yes	Yes

Notes: The table shows the change in citations that scientists received after the year in which the stars with whom they collaborated were either nominated for or won the Nobel Prize (the "event year"). The dependent variable is the natural logarithm of one plus the number of citations that an article accumulated up to a given year (excluding any self-citations). Control variables include *Article age* and *Career age*. Robust standard errors are clustered at the article-level and reported in parentheses. $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.10$.

Table A.5: Effect of collaboration on citations – Alternative control groups

– *"Best" nominees* –

Dependent variable	*Log cumulative citations*							
Nominee sub-sample	*Best nominee by number of nominations*		*Best nominee by number of nominators*		*Best nominee by aggregated authority of nominators*		*Best nominee by number of Nobel Laureate nominators*	
Article sub-sample	Citations to articles published prior to Nobel Prize award	Citations to articles published prior to Nobel Prize award & first collaboration	Citations to articles published prior to Nobel Prize award	Citations to articles published prior to Nobel Prize award & first collaboration	Citations to articles published prior to Nobel Prize award	Citations to articles published prior to Nobel Prize award & first collaboration	Citations to articles published prior to Nobel Prize award	Citations to articles published prior to Nobel Prize award & first collaboration
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post award	0.013***	0.009***	0.014***	0.009***	0.004*	0.001	0.005**	0.004*
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)
Scientist collaborated with winner × Post award	0.010***	0.010**	0.011***	0.011**	0.021***	0.019***	0.021***	0.017***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)
Observations	1,224,194	653,121	1,210,515	647,356	1,215,199	639,357	1,436,515	759,151
Number of clusters	24,536	10,909	24,307	10,837	21,434	9,688	25,245	11,611
R-squared	0.806	0.783	0.807	0.785	0.801	0.777	0.796	0.767
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Article FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field of study × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows the change in citations that scientists received after the year in which the stars with whom they collaborated were either nominated for or won the Nobel Prize (the "event year"). The dependent variable is the natural logarithm of one plus the number of citations that an article accumulated up to a given year (excluding any self-citations). Control variables include *Article age* and *Career age*. Robust standard errors are clustered at the article-level and reported in parentheses. $^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.10$.