

SOCIAL SCIENCES

Prepublication disclosure of scientific results: Norms, competition, and commercial orientation

Jerry G. Thursby,^{1,2} Carolin Haeussler,³ Marie C. Thursby,^{1,4,5,6*} Lin Jiang^{7,8}

On the basis of a survey of 7103 active faculty researchers in nine fields, we examine the extent to which scientists disclose prepublication results, and when they do, why? Except in two fields, more scientists disclose results before publication than not, but there is significant variation in their reasons to disclose, in the frequency of such disclosure, and in withholding crucial results when making public presentations. They disclose results for feedback and credit and to attract collaborators. Particularly in formulaic fields, scientists disclose to attract new researchers to the field independent of collaboration and to deter others from working on their exact problem. A probability model shows that 70% of field variation in disclosure is related to differences in respondent beliefs about norms, competition, and commercialization. Our results suggest new research directions—for example, do the problems addressed or the methods of scientific production themselves shape norms and competition? Are the levels we observe optimal or simply path-dependent? What is the interplay of norms, competition, and commercialization in disclosure and the progress of science?

INTRODUCTION

Scientific progress and open, timely disclosure of results go hand in hand. In principle, the Mertonian norm that scientific discoveries belong to the community, coupled with the priority rule that credit belongs to the first to discover a result, ensures open and timely disclosure of results (1–3). Open disclosure enables researchers to build on each other's work, minimizes duplication of effort, and facilitates collaboration, feedback, and verification (4, 5). However, before publication, scientists who disclose their results risk being scooped by competitors in the race to publish. Those whose research has commercial potential may have incentives to keep results from the public domain before patenting, and researchers supported by industry may be required to delay or not publish certain results. We see the tension between prepublication disclosure and secrecy in discussions of preprints in the biological sciences, particularly in the more competitive areas (6, 7). Anecdotally, prepublication disclosure is avoided in some subfields of physics, a field in which posting preprints has been a norm for decades (for example, ArXiv), and in computer sciences, researchers have been known to disclose results only on the day of journal submission (3). With a time stamp, preprint servers can establish priority (albeit not validation) in fields with long lags in peer-reviewed publication (8, 9).

Although there is extensive research and debate on sharing academic science in general (10–21), we know little about open disclosure before publication. There is a rich tradition of survey research on the failure of scientists to respond to requests for data or materials (10–14) and on their reluctance to discuss current research for fear of being scooped (12, 15, 16). Neither strand of research has examined the question of open disclosure. Materials exchange is not disclosure because requests for materials presume knowledge of their existence. Regarding discomfort in discussing research, it is not clear whether the discussion is with one or many scientists. An exception is the study of Haeussler

et al. (17), which shows that the incentives and trade-offs for sharing one-to-one information or material sharing are quite different from those motivating general, public presentation of results before publication. Whereas expectations of reciprocity govern the former, concerns over misappropriation and potential feedback govern the latter. There is research on “disclosure,” but the term is taken to mean publication. Papers in this stream either study withholding or publication delays (12, 18–20) or theoretically model the researcher's dilemma on when to publish (21). Thus, research on prepublication open disclosure is neglected, leaving the public debates largely uninformed.

Filling this gap is our interest. We ask, given the risks of being scooped, why do those who disclose do so? Do they tend to withhold crucial results so that the disclosure is more apparent than real? How is disclosure related to individual characteristics and research profiles? How is it related to perceived norms and competition, or commercial orientation? Further, although the bulk of previous research has focused largely on life sciences [with select work on secrecy in mathematics and physics (15, 16)], we examine open disclosure across a wide array of fields.

Our analysis is based on survey responses and publication data for 7103 active faculty researchers in the United States, Germany, and Switzerland across nine fields as measured by department/school affiliation. Collectively, these researchers have 136,815 publications over the period 2010–2014, which had received 1,217,437 citations by 2017. Department or school affiliations are agriculture, biological science, computer science, engineering, mathematics, medical school basic science, medical school clinical, physical science, and social science. Survey details and analysis of response bias are in Methods.

We make three contributions to the literature on scientific disclosure before publication. First, for a publication in an area considered important to the respondent's research portfolio, we report whether the respondent openly disclosed the results before publication and, if so, the motivation and stage of the research when disclosed. Confirming our initial hypotheses, we find that the fear of being scooped (an element of competition) is related to whether the results were disclosed prepublication and, if disclosed, at what stage. Among the reasons for disclosing, we examine obtaining credit before another group could disclose a similar result because the fear of being scooped may prompt disclosure rather than waiting for publication. We also examine feedback, attracting collaborators, attracting others to work on

Copyright © 2018
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Laboratory for Innovation Science at Harvard, Cambridge, MA 02138, USA.
²TyGeron Institute, Nashville, TN 37220, USA. ³University of Passau, 94032 Passau, Germany. ⁴National Bureau of Economic Research, Cambridge, MA 02138, USA.
⁵Vanderbilt University, Nashville, TN 37203, USA. ⁶Georgia Institute of Technology, Atlanta, GA 30308, USA. ⁷University of Missouri, Columbia, MO 65218, USA. ⁸University of South Florida, Tampa, FL 33620, USA.

*Corresponding author. Email: marie.thursby@scheller.gatech.edu

their research problem, and deterring others from working on their part of the problem. Across all fields, obtaining feedback is the most important and deterrence is the least important, but there is substantial field variation among the other motivations, with the more formulaic fields disclosing more for credit, as well as attracting others to work on the problem, while at the same time attempting to deter others from their part of the problem.

Second, we construct ordinal indices of researcher disclosure “types,” which reflect how often respondents present unpublished or unpatented results to general audiences and how often crucial parts are withheld when presenting. We define disclosure to general audiences to include web postings, preprints, and conferences. By including the frequency of withholding critical results, the index captures the fact that scientists avoid disclosing not only by not presenting but also by withholding in presentation. This inclusion for prepublication presentation is in the same spirit as Blumenthal *et al.*'s attention to withholding in publications (10, 12). As in the results on disclosure of a particular paper, we find substantial variation across fields.

Third, on the basis of a probability model and series of sensitivity exercises, we show that 70% of the field differences in disclosure types are correlated with (that is, statistically explained by) variables reflecting norms, competition, and commercial orientation, which we collectively call the NCC variables. These three sets of variables embody the tension between the Mertonian norm, with its expected positive impact on disclosure, and negative incentives presented by competition and commercial potential. Further, we show that the observed field variation in disclosure is not related to marginal (partial) effects of the NCC variables; instead, field differences are related to levels of the NCC variables across fields. The model includes a variety of measures of respondent perceptions of norms, competition, commercial orientation, as well as demographics such as age and rank, productivity measures such as publications and citations, characteristics of research profile such as basic research or interdisciplinary direction, size of the research group, and field fixed effects.

The analysis stands in marked contrast to previous work not only in relating the NCC variables to open disclosure prepublication but also in capturing multiple ways in which researchers perceive norms and competition or are engaged in commercial activities. In terms of norms, we introduce variables to reflect respondent perceptions of the extent to which researchers in their field openly disclose results before publication, whether they receive valuable feedback when presenting to general audiences, and a measure of the norm-based system of protecting intellectual property by which researchers are expected to acknowledge authors when using their results (22, 23). Regarding competition, scientists who present before publication risk providing competitors with information to scoop them in the race to publish, a loss even with proper acknowledgement. Whereas previous studies have measured competition only by indices centered on the fear of being scooped (15, 16), we use a number of measures of competition. One such measure is the respondent perceptions of competition to publish. Although influencing the public is an important goal for some scientists, publishing remains the holy grail for many, if not all (24, 25). We also include the number of other researchers working on the same problem as the respondent in line with the economic notion of competition. Notably, we include respondent perceptions of the priority rule in science, the esteem associated with being the first to discover a result (3, 26). In addition, we include respondent perceptions of the level of competition for government funding and students, and the extent to which they choose topics based on the likelihood of funding. Regarding commercial orientation, in ad-

dition to the commonly used measures (patents, startups, and industry funding), we include an additional, and ex ante, commercialization variable to reflect the extent to which respondents choose their research for its commercial potential.

RESULTS

Prepublication disclosure: When? Why?

We asked respondents to think about the problems currently researched in their group. Among those, they were to identify one overarching problem they view as particularly important and to select one publication coming out of their research on the problem. The results in this section are with reference to that publication.

Consider first whether respondents disclosed results from the project to a general audience (for example, web posting, preprint, or conferences) before publication. Figure 1A shows by field the stages of first disclosure: conceptual, after being sure of the validity of the results, completion of a draft or submission, and after publication. Overall, 67.2% of the respondents reported disclosing before publication. Only 6.35% disclosed to a general audience at the concept stage. Roughly 40% disclosed after they were sure of the validity, and 21.4% disclosed once there was a paper, either in draft form or after submission. This leaves roughly one-third who did not disclose prepublication. There is substantial field variation, with the computer scientists and engineers least likely to disclose prepublication. Social scientists are the most likely to disclose at the conceptual stage.

Figure 1B relates the stage of disclosure to respondent concerns over whether the disclosure might lead a competitor to complete a larger research question before them. As the level of concern increases, the percent disclosing at concept or validity falls, whereas the fraction who did not disclose prepublication rises. Relatively few of those who disclosed before publication reported being quite concerned. Roughly 70% of those who said they were very concerned did not disclose until publication. More than half of those who were not concerned at all disclosed at the concept stage or once they were sure of validity.

For those who disclosed results prepublication, we asked the importance (coded 1 = not important to 5 = very important) for each of five, non-mutually exclusive objectives: to obtain feedback, obtain credit before another group can disclose a similar result, attract collaborators, attract others to work on the problem, and deter others from working on this part of the problem. Although attracting others to work on a problem and attracting collaborators might be considered as the same for some respondents, in our presurvey interviews, several interviewees noted that they disclose primarily to attract others to work in their area, though not as collaborators. Computer scientists said there was a reputational benefit to starting new research directions; both they and mathematicians mentioned wanting others to work on problems they could not solve. Many respondents viewed these reasons as distinct because 48.4% of respondents gave different answers on attracting collaborators and attracting others to work in the field.

Figure 1C gives the importance of objectives by field; it shows average responses by field. There is little variation in the importance of feedback by field, with feedback being the most important motive. For feedback, only agriculture is significantly different from other fields. For deterrence, mathematics is significantly different from all fields except computer science and medical school basic. Obtaining credit is most important for computer science, engineering, mathematics, and physical sciences—notably the most formulaic fields. Obtaining credit is least important for agriculture, which is significantly different from all

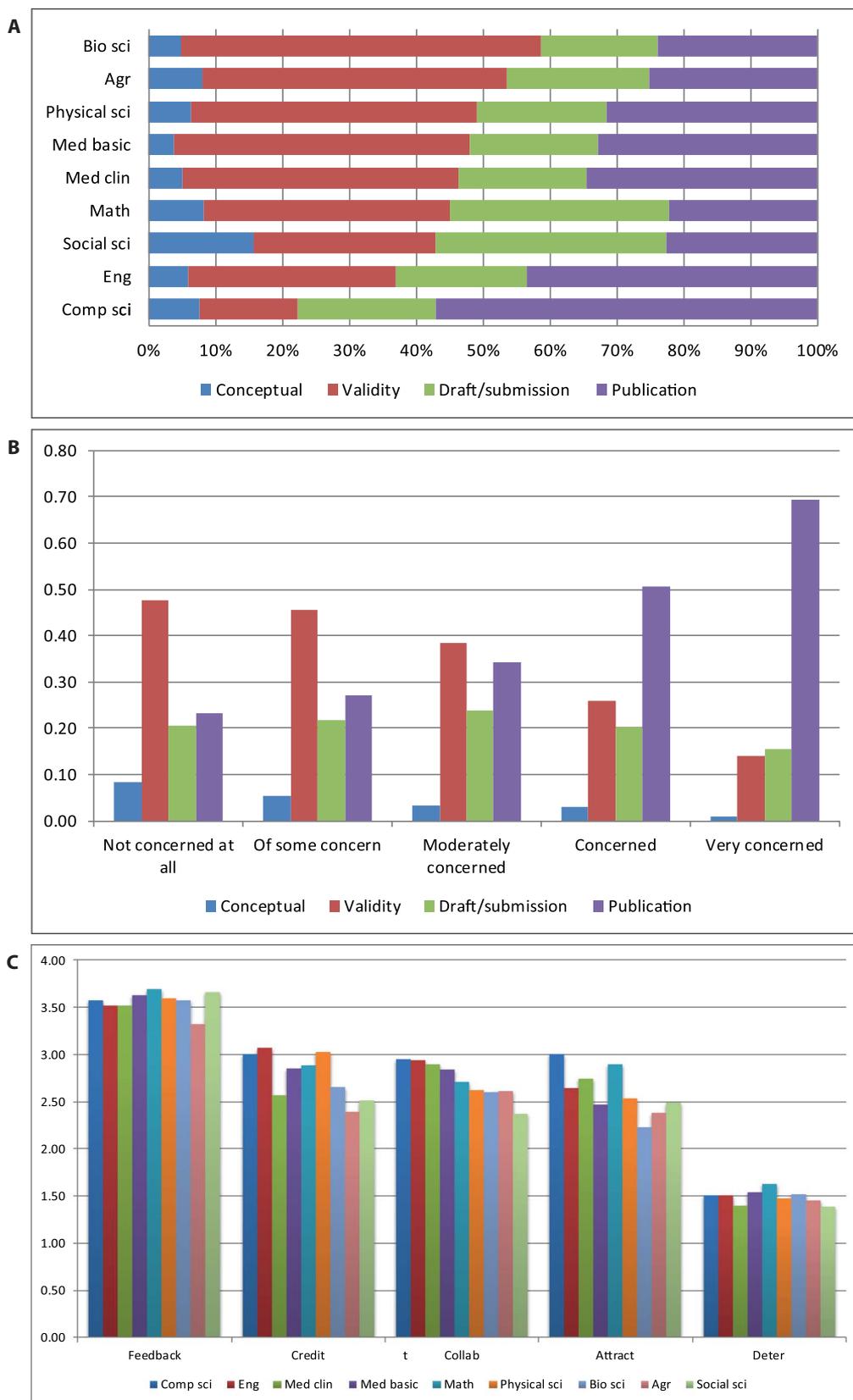


Fig. 1. Disclosure of an identified paper. (A) Stages of disclosure to general audience by field. **(B)** Stages of disclosure and concern of being scooped. **(C)** Motivation for disclosure by field.

other fields except social sciences. Attracting collaborators is least important for social scientists and most important in computer science, engineering, and both clinical and basic medical school research. Attracting others to work on a problem has the greatest variation across fields. Computer scientists and mathematicians (not statistically different at the 5% level) attach the most importance to this reason, and biological sciences the least.

One seemingly contradictory observation is that computer scientists and mathematicians are higher than other fields in both attracting others to work on “this problem” and deterring others from working on their “part” of the problem. Further, the same respondents can be high in both. Slightly less than 19% of computer scientists and mathematicians responded that deterrence was moderately to very important, and 70% of these respondents indicated that attracting others was moderately to very important. Informal discussions with researchers in these fields suggest that the reason follows from the formulaic nature of both computer science and mathematics. The formulaic nature allows researchers to more concretely communicate their exact part of the problem and to more easily differentiate their part of the problem from other parts.

Disclosing types

There are two ways to not disclose results: not to present them and, when presenting, to withhold crucial parts. We create an ordinal index of disclosing type that reflects both approaches. Our index uses respondent answers to two questions: (i) how often they present to general audiences and (ii) how often they withhold crucial parts when presenting. Table S1 (parts a and b) shows responses.

Non-disclosers or non-disclosing (ND) type are defined as researchers who report that they never or rarely present prepublication material to general colleagues and/or, when presenting, often or very often withhold crucial parts. Disclosers or disclosing (D) types are identified by responses that they often or very often present and never or rarely withhold crucial parts. D types must satisfy two conditions, whereas ND types must satisfy only one of two conditions. Ambivalent disclosing (AD) types consist of all others. We examine validity of this classification in Methods.

The ND group is 23.9% of the sample, AD is 38.9%, and D is 37.2% of respondents. In the population, ND types may be less likely to respond to the survey, and we find that ND types are less likely to complete the survey once started than D types. This implies the possibility that ND types in the population are more prevalent than in our sample. Response bias is considered in Methods.

As shown in Fig. 2, the percentages above mask considerable field heterogeneity. ND varies from 11.8% in social sciences to 35.3% in engineering. Biological science, which has attracted most attention, is not an extreme by our measure. Except for engineering and computer science, D exceeds ND, with the ratio greater than two for mathematics and social sciences. As shown in section S1 and fig. S1, significant field heterogeneity also extends to responses across the 27 Scopus publication fields in which respondents published. In the next sections, we focus on the heterogeneity in disclosing type across fields as determined by department/school affiliation.

Probability of disclosing

We use an ordered logit model relating the probability of respondent type (ND, AD, or D coded 0, 1, or 2, respectively) to variables thought to be related to disclosing and field fixed effects to account for unobserved heterogeneity. The variables considered are defined in Table 1,

with summary statistics in table S2 and additional details of variable construction in section S2. Fourteen of the 39 independent variables included are NCC variables. Together, the NCC variables reflect the tension of the norms of science (N) with competition (C) and commercial orientation (C). In the theoretical model of Haeussler *et al.* (17), an increase in a scientist’s confidence that others will acknowledge results and other scientific norms should lead to an increase in general disclosure. By contrast, any of the variables reflecting increased competition are expected to decrease the likelihood of general disclosure (15–17). Similarly, we expect academic researchers whose research is commercially oriented to be less likely to openly disclose the results before publication or patenting (15–17).

Table 1A gives the results, expressed as odds ratios, for the ordered logit when all 39 variables are included. Table 1B gives the results after excluding variables with very small *t* statistics (<1). We consider this model as our base model and discuss all results with respect to this model. It is noteworthy that respondent publications and citations as well as measures of the size of their research group have small *t* statistics and thus are dropped. The only regressors of the reduced model not significantly different from zero at a 10% level or less are *Male* and *SumResAreas*. Because the age of the respondent enters both as the level and as the square, we jointly tested *Age* and *AgeSq* and accepted the null hypothesis that they are jointly not different from zero ($P = 0.202$). All else equal, the more laboratory members or collaborators working with a scientist, the less the need for feedback from the external community (17). Variables in the regression that could reflect such feedback possibilities are the number of groups worldwide in which respondents had a trusted colleague (*GrpWWTrust*) and the number of these in which they had a collaborator (*GrpWWCollab*). Whereas the former increases disclosure, the latter is found to decrease it.

The model should not be viewed as causal, but as descriptive, providing partial correlations of variables thought to be associated with disclosure. However, the issue of endogeneity is worth discussing. Of the variables of interest, those most likely to suffer from endogeneity are respondent perceptions of norms; for example, respondent perceptions of the norms of others in their field are likely to be affected by their own norms. With this in mind, we examined endogeneity of norms; details are found in section S2. With reservations detailed in the Supplementary Materials, we suggest that endogeneity of norms can be discounted.

The NCC variables (with one exception) have odds ratios in line with our expectations. Regarding norms, we would expect a belief that

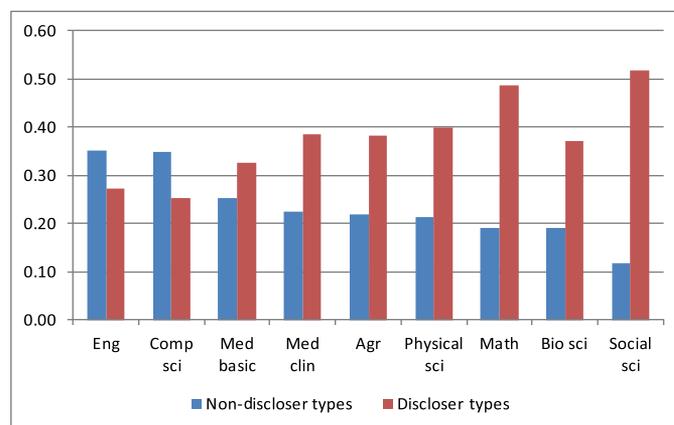


Fig. 2. Fraction of disclosing and non-disclosing types by field.

Table 1. Ordered logit results. *OpenExch*, open exchange is practiced in my field (1 to 5, 5 = strongly agree); *Feedback*, disclosing leads to valuable feedback (1 to 5, 5 = strongly agree); *Acknowl*, researchers use the results of others without acknowledgement (1 to 5, 5 = strongly disagree); *ResCommer*, choose research topics with commercial potential (1 to 5, 5 = strongly agree); *LStartup*, log of number of startups with respondent as founder or member, scientific advisory group; *LPatentAp*, log of number of patent applications; *DIndustry*, indicator equal to 1 if respondent has industry-funded research in the past 3 years; *CompOverall*, perception of overall competition (1 to 5, 5 = very competitive); *CompPub*, perception of competition to publish or present new results (1 to 5, 5 = very competitive); *CompGovFnd*, perception of competition for government research funding (1 to 5, 5 = very competitive); *CompStudents*, perception of competition for students or post docs (1 to 5, 5 = very competitive); *ResFunded*, choose research depending on how likely to get funded (1 to 5, 5 = strongly agree); *HighEsteem*, first to come up with a result is highly esteemed by peers (1 to 5, 5 = strongly agree); *GrpWW*, number of competing groups worldwide (capped at 51); *GrpWWTrust*, number of competing groups worldwide with trusted colleagues; *GrpWWCollab*, number of competing groups worldwide with collaborators; *DetectEasy*, easy to detect failure to acknowledge (1 to 5, 5 = strongly agree); *ResLeader*, my research group is considered to be among the leaders in the field (1 to 5, 5 = strongly agree); *Lpubs*, log of number of publications in the past 5 years; *Lcites*, log of number of citations in the past 5 years; *NSF_DFG*, indicator = 1 if NSF or DFG funding; *DFG*, Deutsche Forschungsgemeinschaft, the German Research Foundation; *NIH_EC*, indicator = 1 if NIH or EC funding; *EC*, European Commission; *OtherGovFnd*, indicator = 1 if other government funding; *OtherFnd*, indicator = 1 if funding other than government or industry; *ResBasic*, my research is basic (1 to 5, 5 = strongly agree); *ResInterd*, my research area is interdisciplinary (1 to 5, 5 = strongly agree); *ResRealW*, my research is driven by real-world problems, with or without commercial potential (1 to 5, 5 = strongly agree); *ResReput*, my reputation among academics is important to me (1 to 5, 5 = strongly agree); *Assistant*, indicator = 1 if assistant professor; *Associate*, indicator = 1 if associate professor; *Male*, indicator = 1 if male; *Age*, respondent age; *AgeSq*, square of age; *NumGroup*, number of full-time researchers in research group; *NumReport*, number of researchers who report directly to respondent; *Migrant*, indicator = 1 if not working in country of birth; *BirthDevelop*, indicator = 1 if birth country developed.

		Part a			Part b		
		Odds ratio	t statistic	P	Odds ratio	t statistic	P
Norms	<i>OpenExch</i>	1.108	2.79	0.005	1.103	2.76	0.006
	<i>Feedback</i>	1.248	5.14	0.000	1.250	5.35	0.000
	<i>Acknowl</i>	1.095	3.22	0.001	1.098	3.43	0.001
Commercial	<i>ResCommer</i>	0.779	-8.13	0.000	0.779	-8.4	0.000
	<i>LStartup</i>	0.890	-1.37	0.170	0.869	-1.70	0.089
	<i>LPatentAp</i>	0.762	-5.48	0.000	0.770	-5.46	0.000
	<i>DIndustry</i>	0.760	-3.39	0.001	0.779	-3.20	0.001
Competition	<i>CompOverall*</i>	1.007	0.14	0.887			
	<i>CompPub</i>	0.845	-3.80	0.000	0.851	-4.88	0.000
	<i>CompGovFnd</i>	1.070	1.57	0.116	1.084	2.06	0.039
	<i>CompStudents*</i>	1.020	0.58	0.559			
	<i>ResFunded</i>	0.895	-3.57	0.000	0.889	-3.96	0.000
	<i>HighEsteem</i>	0.872	-3.29	0.001	0.876	-3.31	0.001
	<i>GrpWW</i>	0.992	-3.79	0.000	0.993	-3.66	0.000
Other	<i>GrpWWTrust</i>	1.044	7.18	0.000	1.046	7.51	0.000
	<i>GrpWWCollab</i>	0.973	-2.42	0.015	0.970	-2.77	0.006
	<i>DetectEasy*</i>	1.000	0.01	0.995			
	<i>ResLeader</i>	1.115	3.07	0.002	1.096	2.83	0.005
	<i>Lpubs*</i>	1.014	0.20	0.845			
	<i>Lcites*</i>	0.987	-0.33	0.739			
	<i>NSF_DFG</i>	1.197	2.58	0.01	1.198	2.71	0.007
	<i>NIH_EC*</i>	0.965	-0.47	0.638			
	<i>OtherGovFnd</i>	0.862	-2.21	0.027	0.888	-1.84	0.066
	<i>OtherFnd*</i>	0.995	-0.08	0.936			
	<i>ResBasic</i>	1.042	1.54	0.123	1.047	1.82	0.068
	<i>ResInterd*</i>	1.008	0.22	0.826			

continued on next page

	Part a			Part b		
	Odds ratio	t statistic	P	Odds ratio	t statistic	P
<i>ResRealW</i>	1.056	1.83	0.067	1.061	2.1	0.036
<i>ResReput*</i>	0.998	−0.04	0.965			
<i>Assistant</i>	0.685	−2.96	0.003	0.678	−3.27	0.001
<i>Associate</i>	0.869	−1.57	0.117	0.842	−2.03	0.043
<i>Male</i>	1.079	1.06	0.291	1.051	0.71	0.475
<i>Age</i>	0.964	−1.40	0.162	0.958	−1.66	0.096
<i>AgeSq</i>	1.000	1.47	0.142	1.000	1.74	0.081
<i>NumGroup*</i>	1.002	0.65	0.515			
<i>NumReport*</i>	1.002	0.52	0.606			
<i>Migrant</i>	0.860	−1.86	0.063	0.859	−2.27	0.023
<i>BirthDevelop*</i>	1.058	0.48	0.628			
<i>SumResAreas</i>	1.035	1.34	0.181	1.032	1.42	0.155
<i>USSample</i>	1.710	4.91	0.000	1.693	5.05	0.000
Field fixed effects	Yes			Yes		
Number obs	4247			4445		
Pseudo R-sq	0.084			0.082		

*Variable is dropped from original regression due to t statistic less than 1 in absolute value.

others follow the norms to be accompanied by a greater willingness to openly disclose information before publication. The odds ratios for open exchange, feedback, and acknowledgement are 1.103, 1.250, and 1.098, respectively. The aggregate odds ratio is 1.514 (the product of the individual odds ratios); an increase by one unit in each measure of norms is associated with a 0.514 increase in the odds of D relative to ND or AD. In contrast, a competitive research environment is expected to be associated with less disclosure. For competition variables, the odds ratios for esteem, competition in publishing, number of competitors (*GrpWW*), competition for government funding, and choice of research topics based on likelihood of funding are 0.876, 0.851, 0.993, 1.084, and 0.889, respectively; the aggregate is 0.714. These results also confirm expectations that competition reduces disclosure, except for competition for government funding. The NCC variables most studied in previous work on one-to-one disclosing, those related to commercial orientation, are also expected to reduce prepublication disclosure. The odds ratios for patent applications, startups, industry funding, and research choice with commercial potential are 0.770, 0.869, 0.779, and 0.779, respectively; the aggregate is 0.406. Notably, the absolute value of the norms and competition variables are not statistically significantly different, that is, a unit increase in each of the norms variables is offset by a unit increase in each of the competition variables. The absolute value of the commercial effect is statistically significantly larger than that of the norms or competition variables.

With regard to field differences, two points are important. First, we tested for equivalence of each pair of field fixed effects (36 tests). The fixed effect for computer science is significantly different (5% level) from every other fixed effect, and in every case, there is less prepublication

disclosure by computer scientists. Further, with the exception of engineering versus medical clinical (engineers disclose less: $P = 0.036$), in no other comparison is there a statistically significant difference in fixed effects. The 27 other regressors capture all field heterogeneity with the exception of computer science (and the one case noted above) for which there remains unobserved heterogeneity.

Second, the logit model assumes equal partial effects across fields. We tested this assumption for each of the 12 NCC variables—a total of 432 pairwise tests (for each pair of fields, we conduct 12 tests). Using a 10% (5%) level of significance, we find that 11.34% (6.25%) of the tests are significant. If the null of no difference is correct, then we expect 10% (5%) of the comparisons to be significant. The difference between 10% (5%) and 11.34% (6.25%) is not statistically significant ($P = 0.156$ and 0.0996).

Whereas our measures of ND, AD, and D, as well as the NCC variables, are specific to the individual, in the next section, we show that individuals tend to be more similar to others in their field than to those in other fields with respect to the tension of norms with competition and commercial orientation. Further, collectively, the NCC levels by field statistically explain most of the variation in disclosing by fields. Finally, field differences in behavior arising from the NCC variables are related to differences in the levels of the variables instead of differences in responses to the variables (that is, responses captured by regression coefficients; see the pairwise coefficient test results noted above).

Sensitivity analysis: The role of the NCC in field differences

To explore the role of the NCC variables in disclosing, we conduct two sets of sensitivity analyses. Each analysis is based on the estimated

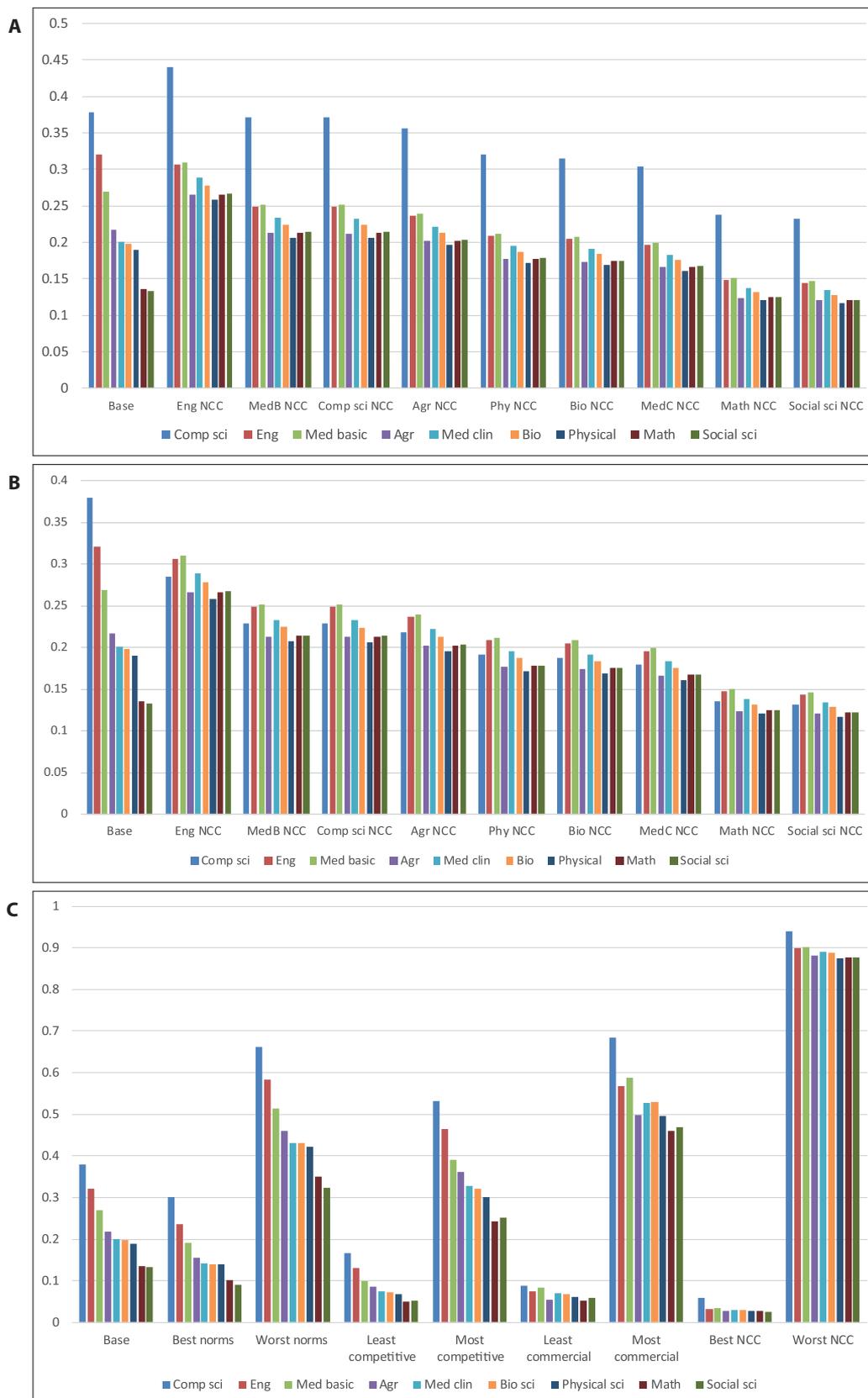


Fig. 3. Sensitivity analysis: Field differences. (A) Predicted probabilities of non-disclosing types: Base and hybrids. (B) Predicted probabilities of non-disclosing types: Base and hybrids with computer science adjustment. (C) Predicted probabilities of non-disclosing types: NCC most and least conducive to disclosing.

coefficients of our probability model (Table 1B). To provide a base of reference for all of the sensitivity analyses, we compute the predicted probability (pp) of being an ND for each respondent using the estimated coefficients. Additional details are provided in the “Statistical methods” section. The average pp by field is presented in the panels labeled “Base” of Fig. 3 (A to C).

Field effects and NCC levels

In the first sensitivity analysis, we compute the average value of each NCC variable by field. For each respondent in the sample, we replace their individual NCC values with the average values of one of the fields—we refer to these as “hybrid” individuals—and compute the pp of ND for the hybrids. We then compute the average pp for each hybrid field. For example, in one of the exercises, the average NCC values for engineering replace each respondent’s actual NCC values. Our hybrids all have the same NCC values (in this example, it is the average engineer’s NCC), but they retain the original values of all other variables. The exercise is repeated using each field’s average NCC.

This exercise has 81 different results; for each of the nine sets of average NCC values, we have the average hybrid pp for each of the nine fields. Figure 3A presents results. For example, the second panel of Fig. 3A (labeled “Eng NCC”) shows the average values of the pp of ND for each field, where every individual has the NCC values of the average engineer. Likewise, the third panel (“MedB NCC”) shows the average pp when each respondent is given the average NCC of the medical basic science respondents. Depending on which field’s NCC averages are used, disclosing behavior is different. For example, excluding computer science, the pp of ND for hybrids using engineering NCC averages varies from 0.258 to 0.306, whereas hybrids using math have pp values from 0.121 to 0.151.

Two features of Fig. 3A are striking. First, consistent with the fixed-effects comparisons, computer science is an outlier. Regardless of the NCC values used, the pp of the hybrid computer scientist is largest. Computer scientists interviewed put forward the hypothesis that the fixed effect comes from the fact that conference proceedings are their preeminent publication outlet. Journal publications are often revised, extended results or are reserved for reviews or compendia (27, 28).

For our computer science respondents, 68.6% of publications in the previous 5 years were conference proceedings, compared to 36.3% for engineering and less than 12% in other fields. To examine this hypothesis on conference proceedings, we replaced the dependent variable in the probability model first with responses to the general disclosing question and second with responses to the withholding question. Computer science continues to be an outlier in the first regression but not in the second. This is consistent with the computer science fixed effect resulting from the role of conferences. If we set the fixed field effect for computer science to the average level of the other fields, then computer science is no longer an outlier (see Fig. 3B).

Second, there are substantial cross-panel differences, but within panels, the differences are small (except for computer science). The within and cross-panel results suggest the importance of NCC values because across panels, the NCC values differ, whereas within a panel, the NCC values are the same. Note that all within-panel differences in pp across fields are less than in the Base panel. The median difference among the base predicted probabilities is 0.0725, and for the hybrids, it is 0.0224. This is a reduction of 69.9%. Excluding computer sciences, the median reduction is 78.9%.

Relative effects of NCC variables

To examine the relative effects of norms, competition, and commercial orientation on disclosing, we conduct a second set of sensitivity exer-

cises. Each respondent is successively given norm values that are most and then least conducive to disclosing. For example, the norms variables *OpenExch*, *Feedback*, and *Acknowl* are changed from actual (observed) values to 5, which is the value of each variable that is most conducive to disclosing. In the other norm exercise, we give each norm variable the value 1, the value that is least conducive to disclosing. All other NCC variables retain their observed values. This is followed by exercises in which each respondent is given competition variable values most and then least conducive to disclosing and commercial variable values most and then least conducive to disclosing.

The results are shown in Fig. 3C. For comparison, we repeat in the first panel the base probabilities from Fig. 3A. The second panel of Fig. 3C report results when the norm values are most conducive to disclosing (“Best Norms”). This is followed by an exercise in which we use values of norms that are “worst” for disclosing (“Worst Norms”). This process is continued for competition and commercialization (see the “Statistical methods” section for details on values). In a final exercise, we use all NCC values most and then least conducive to disclosing; results are in the right-most panels of Fig. 3C.

Moving between least and most commercial has the largest effects on disclosing in comparison with norms and competition. If all respondents were the most commercial, then five of the nine fields would have more than 50% ND types and the other four fields would be close. This is quite different from the worst norms and most competitive regimes. Further, the least commercial orientation is closest to the “Best NCC” panel. Having the best norms shows the least difference from the observed base case. For both the least competitive and the least commercial, all but two cases (computer science and engineering in the least competitive case) have fewer than 10% ND types.

Finally, there is very little variation across fields when all have the same set of values for the NCC; this finding mirrors the effects noted in the first set of sensitivity exercises.

Norms, competition, and commercialization

Field differences in disclosing behavior are largely correlated with differences in the 12 NCC variables. Here, we present the levels of those variables by field. To make comparisons, each variable is standardized to have mean zero and variance one. We then extract the values for the norms, commercial orientation, and competition variables by field. The averages are presented in Fig. 4 (A to C). Figures S2 and S3 show the tests of differences across fields. What stands out in each is the diversity across fields.

Figure 4A shows the standardized values for the norms variables. The highest level of each norm is observed in mathematics followed by physical sciences. The lowest levels are found in medical school basic sciences. These results are consistent with the result in Fig. 2 that mathematics and physical sciences have a significantly greater degree of disclosure to general audiences than medical basic sciences.

Figure 4B shows the results by field for the commercialization variables. The most commercial area is engineering, which is highest in commercial orientation for all measures except the number of startups (highest for computer science). The least commercial areas are social science and mathematics. This is again consistent with the results regarding disclosing and non-disclosing types in engineering, social sciences, and mathematics shown in Fig. 2.

Figure 4C shows the results for the competition variables. The highest average level of competition is in medical basic sciences, but on examining the individual competition variables, that field is not highest

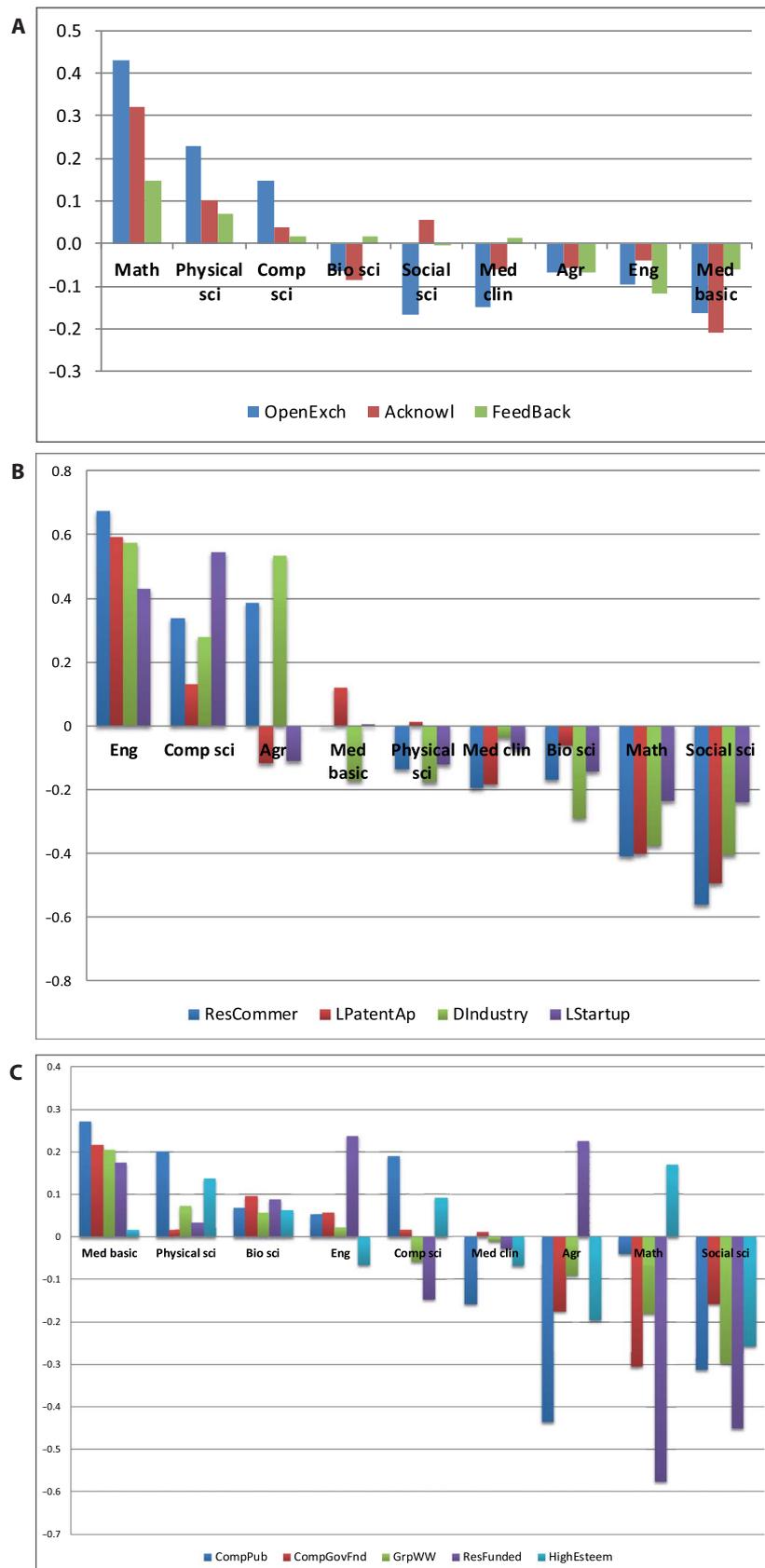


Fig. 4. Standardized NCC values by field. (A) Standardized norms. (B) Standardized commercial orientation. (C) Standardized competition.

for *ResFunded* or *HighEsteem*. The lowest levels of competition are in the social sciences and mathematics (with the exception that mathematics is highest for *HighEsteem*); note that these two fields are also lowest in commercialization. Not coincidentally, social sciences and mathematics have a greater degree of disclosure to general audiences than medical basic science (Fig. 2).

DISCUSSION

According to the Mertonian norm of communalism, scientific results belong to the community at large. This is the rationale for publishing scientific output. More debated is the idea that research should be disclosed before publication. Reasons for prepublication secrecy include the risk of being scooped, endangering patent filing, and, from the perspective of the scientific community, protection from results that would not pass peer review. On the other hand, disclosure of preliminary results may be necessary to find collaborators with complementary skills, early feedback can improve the quality of eventual journal submission, early disclosure can avoid duplication of science, and publication lags suggest that disclosing prepublication may speed the progress of science. Current debates on the pros and cons of prepublication disclosure occur against a backdrop of little systematic research on prepublication disclosure—how frequently it occurs and the factors involved.

We examine prepublication disclosure to general audiences across a wide array of fields and find that, except for engineering and computer science, disclosure is more common than not (as measured both by disclosure of a paper and our index of disclosing types), but there is substantial variation across fields, with social scientists and mathematicians being the most likely to disclose. Notably, the life scientists are not the least likely to disclose. As shown in Fig. 1A, in percentage terms, once they are sure of validity, medical basic and medical clinical scientists are more likely to disclose than mathematicians.

One result of our econometric analysis of disclosing types stands out. It appears that field differences in disclosing do not stem from differences in responses to changes in NCC values; rather, they arise largely from different levels of the NCC variables by field. Further, these level differences account for about 70% of the variation across fields. Note that if more prepublication disclosure is preferred (from a societal standpoint), then our results suggest that the goal might be achieved by concentrating on norms, competition, and commercialization.

As for why one would disclose, obtaining feedback is the most important reason in every field. Although deterrence is the least important, a considerable number of researchers consider deterrence when deciding to disclose prepublication. In the more formulaic fields, scientists often present before publication to attract competitors to work independently on a problem while at the same time wanting to deter work on their part of the problem. This view of competition is in marked contrast to a racing model where scientists fear being scooped by competitors. Moreover, it is the formulaic fields where prepublication disclosure to obtain credit is more prevalent.

The results on attracting others are thought-provoking. Although we can conjecture, we do not know why the respondents want to attract others. The motivation could be similar to that for credit. As scientists in our presurvey interviews suggested, the scientist may want credit for discovering a problem not yet solved—perhaps a new direction for the field so that they might be thought of as a research leader—or the motivation may be similar to that for attracting collaborators (29); the scientist may want scientists with complementary skills to work on the

problem regardless of collaboration (30). More generally, the importance of both “attraction” motives (to collaborate and to work on the problem) suggests that government funding policies meant to encourage collaboration may increase prepublication disclosure. This hypothesis may explain the somewhat anomalous positive correlation of competition for government funding and the probability of disclosure. Is this an unintended consequence of such policies? Are such consequences important to consider in evaluating the policies (31)?

It is important to recognize our results as descriptive rather than prescriptive. We make no welfare statements about prepublication disclosure. There are well-known risks to those who disclose, and Fig. 1B highlights fear of being scooped as a consideration in the disclosure decision. Further, the disclosure of unrefereed results may have undesirable effects. Attracting collaborators with complementary skills, *ceteris paribus*, should increase scientific progress in the attracting field but only if the increased output outweighs collaboration costs. To some, the term “deterrence” might have negative connotations, but recall that the data we present pertain to deterrence from the part of the problem researched by the presenter. This behavior is beneficial if it reduces unnecessary duplication of effort, but we make no judgment. The same can be said of attracting others to the field. Although it is tempting to think of that as socially desirable, examining the optimal size or rate of progress in a field is beyond our scope.

Future research directions include examining the welfare effects of prepublication disclosure. These studies might include relating what is disclosed and stage of disclosure to community benefits. For example, recent field experiments in computational biology (5) have shown significant benefits from disclosure of intermediate results for subsequent reuse by others. Although we reported the stage of disclosure of results such as web postings, preprints, and conferences, we did not explore the extent to which data, materials, or algorithms were disclosed. Another worthy avenue is to examine the extent to which disclosure restrictions or grants of access to use is also warranted (32). This topic is related to an old stream of research by legal scholars on disclosure in patenting, which restricts reuse (33). Our results also suggest that there is still much that we do not know not only about prepublication disclosure but also about the NCC variables we examine. Is there something about the scientific production function in these fields that shapes differences in norms and competition? For example, what is the role of capital intensity? Further, are observed levels optimal for the production of science, or are they due to exogenous reasons and path dependence? What is the interplay among competition, norms, and commercialization in openness and progress of science?

One point we have not addressed is the extent that individuals are attracted to a field based on openness before publication. While it may well be the case that individuals are attracted (or repelled) according to levels of competition and/or commercial orientation, and possibly the norms in a field, we are not convinced that this alters any interpretation of our results. It still remains, regardless of selection into fields, that some fields disclose more than others, and this is highly correlated with the NCC levels of the fields. Selection into fields, however, would affect the levels of the NCC variables.

METHODS

Survey and publication data

We conducted a survey in 2014 of active researchers regarding prepublication disclosing of research results. The survey was email-based. Email addresses were obtained from department and school websites at

121 universities in the United States, Germany, and Switzerland. The survey was sent to 81,406 email addresses, and 5643 (6.9%) bounced back as invalid. The reason, we believe, for the high percent of invalid addresses is that many departments include addresses of graduate students and post docs alongside faculty addresses. Many of those that bounced back were likely students and post docs who had left their institutions between the time we collected the addresses and the time of the survey.

The remaining 75,763 were valid addresses. Of these, 872 were identified as no longer appropriate for the survey primarily because they had retired, they had moved entirely to administration, or we were notified that they had passed away. This left 74,891, of whom 7263 (9.7%) completed the survey and 1803 (2.4%) partially completed the survey. Respondents were assured that their responses would remain confidential. The overall response rate was 12.1%. The response rate for the U.S. sample is 12.5%, and the European rate is 11.2%. These are statistically significantly different ($P = 0.000$).

The survey was intended for active researchers; thus, we exclude from our analysis the 700 respondents who do not self-identify as active researchers. Although the survey included graduate students, post docs, and research scientists, we consider here only the 7103 respondents who hold faculty rank: 19.9% are assistant professors, 24.2% are associate professors, and 55.9% are professors. U.S. respondents are 89.6% of respondents, 7.9% are in Germany, and the remaining 2.4% are in Switzerland.

Publications and citation data

We obtained from Scopus the total number of publications and citations for each respondent; coverage includes conference proceedings, articles in press, and journal publications. We were able to find publication and citation records for 92.6% of respondents. We exclude any whose records could not be found; these include, obviously, those with zero publications, but we also know that some (most likely the majority) are failures to find Scopus records rather than zero publications. Dropping those with zero publications is not considered an important problem because we wish to restrict attention to active researchers. We tested whether the failure to find records is significantly related to a respondent's disclosing type [disclosers (D), ambivalent disclosers (AD), or non-disclosers (ND)], and we did not find a statistically significant relationship ($P = 0.379$). We did, however, find a significant relationship between field and missing Scopus records ($P = 0.000$). The percent with missing records ranges from 3.5% in computer science (followed by engineering at 4.5%) to 13.5% in social sciences (followed by physical sciences at 11.5%).

Field categorization

We categorize respondents in nine department/school affiliations: agriculture, biological sciences, computer science, engineering, mathematics, medical school basic sciences, medical school clinical, physical sciences, and social sciences. Respondents in schools of agriculture who are in agricultural engineering are classified in engineering rather than in agriculture, and those in agricultural economics or agricultural business are classified in social sciences. Statisticians and biostatisticians are classified in mathematics. In most cases, assignment to a field is simple; however, where affiliation was unclear (for example, some respondents are found in Department of Mathematics and Computer Science), we examined key words provided by respondents to describe their research. Nonetheless, there are likely errors that remain. As a check on our division of medical school faculty into basic sciences and clinical departments, we examined responses to a question regarding whether respondent agreed that their work was basic. More than 73% of those

identified with medical school basic science departments agreed or strongly agreed that their work is basic. For those in clinical departments, the fraction was 38.2%.

Tests for response bias

Comparing response by field

We compared the response rate for engineering with that of mathematics. These are chosen because the willingness to disclose by respondent is markedly different in these two fields; thus, we might expect the willingness to take the survey to be different across these fields. After dropping those for which the email bounced back as invalid, there are 16,720 engineers and mathematicians who received the survey, of whom 10,599 are engineers (63.4%) and 6720 are mathematicians (36.6%). The response rate is statistically significantly higher for mathematicians (13.9%) than for engineers (12.5%). The P value for the difference is 0.008. The lower rate for engineers is not surprising given that, among respondents, there is a substantially higher fraction of non-discloser types.

Comparing response in terms of Scopus publications

In another series of tests, we compared publications and citations between respondents and nonrespondents. We used cluster random sampling to draw a sample of 300 nonresponding academics. Clustering was based on the geographic distribution of our respondents. We then attempted to match these with Scopus publication data. Some could not be matched. This occurs, in part, from the fact that not all academics have publications (and our survey is intended for active researchers only). We did not find a match for 30 U.S. academics, 5 German academics, and 1 Swiss academic. We further lost four U.S. and one German observation due to errors in the download file. Two hundred fifty-nine remain in our sample (86% of the original 300). The geographic distribution is 215 U.S., 30 German, and 14 Swiss observations.

Tests of differences in publications and citations do not suggest that respondents are different from nonrespondents. The number of publications by respondent in the 5 years before the survey is as follows: respondent average, 19.9; nonrespondent average, 21.5; P value in test of difference, 0.4439. The number of citations to publications in 5 years before the survey is as follows: respondent average, 179; nonrespondent average, 207; P value in test of difference, 0.1950. The average cites to publications in 5 years before the survey are as follows: respondents, 2.33; nonrespondents, 2.62; P value in test of difference, 0.2609. The average number of coauthors in 5 years before the survey is as follows: respondents, 5.08; nonrespondents, 5.03; P value in test of difference, 0.8995.

Comparing response by early versus late respondents

We tested for potential nonresponse bias by comparing the answers to questions from the first wave of respondents with the last wave of respondents. Our analysis compares the first 10% of respondents to the last 10%. We tested whether the fraction of disclosing types (D, AD, or ND) are statistically significantly different between the early and late disclosers. We also tested whether these two groups differ in terms of our variables for competition, commercialization, and norms. Differences are not statistically different at conventional levels.

Comparing responses of those who started the survey but did not complete to the ones who finished the survey

Comparing these two groups, we found no significant difference in terms of the share of AD types ($P = 0.328$). However, we find weakly significant ($P = 0.096$) that the respondents who did not complete the

survey are more likely ND types. We find a strong significant effect ($P = 0.014$) for the ones who finished to be more likely D types. However, comparing the likelihood of completing a survey across fields does not suggest that, in fields with a large share of non-disclosers, respondents tend to be less likely to complete the survey. For example, engineering, the field with the largest share of ND types, ranges exactly in the middle of all fields in terms of the frequency of completes.

Validation of respondent disclosing type

Respondent disclosing type was determined by self-identification with regards to general disclosure and withholding. As is generally the case with self-identified characteristics, there can be concerns about validity of the measure. We considered validity in three ways. We first compared our classification of disclosing type into D, AD, or ND with whether respondents disclosed prepublication results of an identified recent paper. Second, we compared our classification of respondents into disclosing type with disclosing with trusted colleagues. Finally, we compared disclosing type with disclosure when a specific request is made to respondent for materials, data, etc. In each case, we find a positive association between our classification of disclosing type (D, AD, and ND) and these other disclosing-related situations; details follow.

Identified paper comparison

We asked respondents to identify a recently published result in an important, overarching area of their research and to indicate whether they had openly disclosed the result before publication. In asking respondents to identify specific incidents, inferences from stereotypical answers or received opinions as well as vague opinions are reduced (34, 35). Respondents' answers are more concrete when tied to specific projects or situations. In allowing respondents to decide what incidents they report, respondents are better in recalling situations, and it reduces the bias of researchers to preselect incidences. Because respondents may tend to select projects or situations that are important to them, we may not catch the full heterogeneity of the general population. Slightly more than 86% of D types disclosed the paper prepublication, whereas only 39.2% of ND types did so. AD types are in the middle (67.6% disclosed).

Recall that disclosing type is determined by the responses to two questions: (i) How often they present to general audiences and (ii) how often they withhold crucial parts when presenting. Response to each question is also significantly correlated with concerns of being scooped for the identified paper. The partial correlation of concern with part a (holding constant the response to part b) is -0.13 , and the partial correlation of concern with part b (holding constant the response to part a) is 0.31 ; both are significantly different from zero ($P = 0.0000$). We view these results and those of the prior paragraph as strong support for the validity of our disclosing type measures.

Trusted colleague comparison

We asked respondents about disclosing prepublication results with trusted colleagues. Table S1 (parts c and d) shows the responses to disclosing with and withholding from trusted colleagues. In our presurvey interviews, we found researchers to be comfortable with the idea of trusted colleagues and our definition of a trusted colleague. A trusted colleague is defined as "an individual outside your research group to whom you could provide confidential information and be sure that it is not passed on or used to compete with you." Research group is defined as a "group of individuals performing research and/or development under common supervision and/or mentorship."

Respondents are divided into three disclosing types according to disclosing with and withholding from trusted colleagues. D-T types are respondents who often or very often disclose with trusted colleagues

and who never or rarely withhold crucial parts. ND-T types are those who never or rarely disclose with trusted colleagues and/or often or very often withhold from trusted colleagues. AD-T types consist of all others. ND-T types comprise 7.9% of the sample (compare to 23.9% for ND types for general disclosing), AD-T types are 27.4% of the sample (AD types are 38.9% of the sample), and D-T types are 64.8% of the sample (D types are 37.2% of the sample). Not surprisingly, disclosing with trusted colleagues is much more common than disclosing with general audiences and withholding from trusted colleagues is much less common.

When we examined the intersection of general and trusted disclosing types, we found a positive and statistically significant ($P = 0.000$) association between disclosing types. The simple correlation of individual types is 0.4 (where ND, AD, and D types are coded 0, 1, and 2, respectively; we similarly code trusted types). The crosstab shows that 32.9% are both D and D-T types, 13.6% are both AD and AD-T types, and 5.4% are both ND and ND-T types. Slightly over half the sample (50.9%) exhibits the same disclosing behavior for both general and trusted colleagues. Further, 41.6% of the respondents exhibit more disclosing with trusted than with general colleagues. In general, those who disclose with general colleagues disclose with trusted colleagues, and those who do not disclose tend not to disclose with either group.

Figure S4A shows the stage of disclosing to trusted colleagues by field of the identified paper discussed above. Figure S4B shows ND-T types by field. For comparison, the figure repeats the fraction NS for general disclosing. The fractions of ND and ND-T follow similar patterns; the simple correlation between the field fractions is 0.819 (significantly different from zero at conventional levels).

Requests for materials comparison

Finally, in our survey, we also asked respondents about their recent experiences with requests for research materials or results. This disclosing, which we call specific disclosing because it involves a specific request to the respondent, has been shown to be very different from the general disclosing that is the focus above (17). We asked them to "think back to the most recent request from someone outside your research group." For that request, we asked for the approximate percentage of information or material they had provided. The average percentage of the specific request provided by D types is 88.1%, for AD types it is 80.5%, and for ND types it is 74.3% (each is statistically significantly different from the others with P values less than 0.004). Figure S4C shows the fraction of requests provided by ND and D types across fields.

Statistical methods

We used an ordered logit model relating the probability of respondent type (ND, AD, or D coded 0, 1, or 2, respectively) to variables thought to be related to disclosing as well as field fixed effects to account for unobserved heterogeneity. We then constructed a series of sensitivity exercises to explore the extent to which field differences are due to differences in NCC variables.

The estimated probability model can be used to calculate each respondent's predicted probability of ND, AD, or D. For example, if $P(Y_i = \text{ND})$ is the probability that person i is type ND, then

$$P(Y_i = \text{ND}) = f(\mathbf{X}'_{1i}\beta_1 + \mathbf{X}'_{2i}\beta_2)$$

where the function f is the logistic probability distribution, vector \mathbf{X}_{1i} contains the 12 NCC (norms, competition, and commercialization)

variables, and the vector X_{2i} contains all other variables. The pp of ND for person i is

$$\hat{P}_i(\text{ND}) = f(\mathbf{X}'_{1i}\hat{\beta}_1 + \mathbf{X}'_{2i}\hat{\beta}_2)$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the estimated values of the unknown coefficients.

We can retrieve the probabilities for respondents in some field A and compute that field's average probability of ND that we denote by $\hat{P}_{AA}(\text{ND})$. Note that these probabilities are not the same as the fraction of disclosing types in the raw data (in part due to the fact that the regression is based on the subset of respondents for whom all variables are observed), although the values are very close; for example, in the raw data, the fractions of ND for engineering and biological sciences are 0.353 and 0.190, respectively, and the average pp values are 0.320 and 0.199. As shown in the leftmost panel of Fig. 3A (labeled Base), $\hat{P}_{AA}(\text{ND})$ ranges from 0.133 in social sciences to 0.379 in computer science. The range of D types is 0.511 in social sciences to 0.214 in computer science. There are 36 possible pairs of fields, and the median difference in pp for each pair of fields for the Base case is 0.073 (the mean is 0.095).

Using the predicted probabilities of ND, we conducted the following sensitivity exercise. We selected some field and computed the average value of each of the 12 NCC variables for that field. For each person in the sample, we replaced their individual NCC values with these average values and then recomputed the pp . We refer to these as our hybrid individuals. That is

$$\hat{P}_{iA}(\text{ND}) = f(\overline{\mathbf{X}}'_{1A}\hat{\beta}_1 + \mathbf{X}'_{2i}\hat{\beta}_2)$$

where $\overline{\mathbf{X}}_{1A}$ is the vector of average NCC variables for field A. Note that we used the estimates of β_1 and β_2 reported in Table 1B. From this, we extracted the average predicted probabilities for each field. Thus, we can compute the average NCC values for engineering and use these as the NCC values for every person. Our hybrids all have the same NCC values (the average engineer's NCC), but they retain the original values of all other variables.

For each of the nine sets of average NCC values, we have nine sets of average predicted probabilities. For example, in Fig. 3A, the second panel, labeled "Eng NCC," shows the average value of the pp of ND for each of the nine fields in the sensitivity exercise, where every individual has the NCC values of the average engineer.

To examine the impact of norms versus competition versus commercial orientation, we conducted a second series of sensitivity exercises. Each respondent was successively given norms variable values that are most and then least conducive to disclosing, competition variable values most and then least conducive to disclosing, and commercial variable values most and then least conducive to disclosing. For example, for the norms exercises, *OpenExch*, *Feedback*, and *Acknowl* were changed from actual (observed) values to 5; these are the values most conducive to disclosing. Note that this exercise provides more evidence on disclosing than the odds ratios reported above because it moves each respondent to the extreme values in the sample.

For most competitive, we set *CompPub*, *ResFunded*, and *HighEsteem* each to 5, whereas *CompGovFnd* was set to 1 (because greater competition for government funding leads to more disclosing). *GrpWW* was set to 51, and *GrpWWTrust* and *GrpWWCollab* were based on the ratio of the average of these variables to the average of *GrpWW* (14.3 and 6.7, respectively). For the least competitive case, *CompPub*, *ResFunded*,

and *HighEsteem* were each set to 1, whereas *CompGovFnd* is set to 5. *GrpWW* is set to 0, and *GrpWWTrust* and *GrpWWCollab* were each set to 0. For most commercial, *ResCommer* is set to 5 and *DIndustry* was set to 1. *LStartUp* and *LPatentAp* are set to the 90th percentile of those with startups and patent applications. For least commercial, *ResCommer* was set to 1, and *DIndustry*, *LStartUp*, and *LPatentAp* were each set to 0.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/5/eaar2133/DC1>

Supplementary Text

section S1. Disclosing type and publication fields

section S2. Construction and summary statistics for variables of ordered logit

table S1. Disclosing: Percentage of respondents.

table S2. Summary statistics.

table S3. The practice of science.

table S4. Perceptions of competition.

table S5. Characteristics of research.

table S6. Competing groups worldwide.

fig. S1. Disclosing by field of publication.

fig. S2. Average standardized NCC values.

fig. S3. Tests of differences in fig. S2 averages.

fig. S4. Other types of disclosing.

REFERENCES AND NOTES

1. R. K. Merton, Priorities in scientific discovery: A chapter in the sociology of science. *Am. Sociol. Rev.* **22**, 635–659 (1957).
2. R. K. Merton, *The Sociology of Science: Theoretical & Empirical Investigations* (University of Chicago Press, 1936).
3. P. E. Stephan, The economics of science. *J. Econ. Lit.* **34**, 1199–1235 (1996).
4. D. B. Resnik, Openness versus secrecy in scientific research abstract. *Episteme* **3**, 135–147 (2006).
5. K. J. Boudreau, K. R. Lakhani, "Open" disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Res. Policy* **44**, 4–19 (2015).
6. J. Berg, N. Bhalla, P. E. Bourne, M. Chalfie, D. G. Drubin, J. S. Fraser, C. W. Greider, M. Hendricks, C. Jones, R. Kiley, S. King, M. W. Kirschner, H. M. Krumholz, R. Lehmann, M. Leptin, B. Pulverer, B. Rosenzweig, J. E. Spiro, M. Stebbins, C. Strasser, S. Swaminathan, P. Turner, R. D. Vale, K. VijayRaghavan, C. Wolberger, Preprints for the life sciences. *Science* **352**, 899–901 (2016).
7. J. Kaiser, The preprint dilemma. *Science* **357**, 1344–1349 (2017).
8. K. Powell, Does it take too long to publish research. *Nature* **530**, 148–151 (2016).
9. R. D. Vale, A. A. Hyman, Point of view: Priority of discovery in the life sciences. *eLife* **5**, e16931 (2016).
10. D. Blumenthal, E. G. Campbell, M. S. Anderson, N. Causino, K. S. Louis, Withholding of research results in academic life science: Evidence from a national survey of faculty. *JAMA* **277**, 1224–1228 (1997).
11. J. P. Walsh, C. Cho, W. M. Cohen, View from the bench: Patents and material transfers. *Science* **309**, 2002–2003 (2005).
12. D. Blumenthal, E. G. Campbell, M. Gokhale, R. Yucel, B. Clarridge, S. Hilgartner, N. A. Holtzman, Data withholding in genetics and the other life sciences: Prevalences and predictors. *Acad. Med.* **81**, 137–145 (2006).
13. M. R. Haas, S. Park, To share or not to share? Professional norms, reference groups, and information withholding among life scientists. *Organ. Sci.* **21**, 873–891 (2009).
14. S. Shibayama, J. P. Walsh, Y. Baba, Academic entrepreneurship and exchange of scientific resources: Material transfer in life and materials sciences in Japanese universities. *Am. Sociol. Rev.* **77**, 804–830 (2012).
15. W. O. Hagstrom, Competition in science. *Am. Sociol. Rev.* **39**, 1–18 (1974).
16. W. Hong, J. P. Walsh, For money or glory? Commercialization, competition, and secrecy in the entrepreneurial university. *Sociol. Q.* **50**, 145–171 (2009).
17. C. Haeussler, L. Jiang, J. Thursby, M. Thursby, Specific and general information sharing among competing academic researchers. *Res. Policy* **43**, 465–475 (2014).
18. J. G. Thursby, M. C. Thursby, Knowledge creation and diffusion of public science with intellectual property rights, in *Intellectual Property Rights and Technical Change, Frontiers in Economics*, K. Maskus, Ed. (Elsevier, 2008), vol. 2.
19. D. Czarnitzki, C. Grimpe, A. A. Toole, Delay and secrecy: Does industry sponsorship jeopardize disclosure of academic research. *Ind. Corp. Change* **24**, 251–279 (2015).

20. J. S. Gans, F. E. Murray, S. Stern, Contracting over the disclosure of scientific knowledge: Intellectual property and academic publication. *Res. Policy* **46**, 820–835 (2017).
21. C. Bobtcheff, J. Bolte, T. Mariotti, Researcher's dilemma. *Rev. Econ. Stud.* **84**, 969–1014 (2017).
22. E. Fauchart, E. von Hippel, Norms-based intellectual property systems: The case for French chefs. *Organ. Sci.* **19**, 187–201 (2008).
23. A. K. Rai, Regulating scientific research: Intellectual property rights and the norms of science. *Northwest. Univ. Law Rev.* **94**, 77–152 (1999).
24. D. A. Scheufele, Communicating science in social settings. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14040–14047 (2013).
25. K. L. Milkman, J. Berger, The science of sharing and the sharing of science. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13642–13649 (2014).
26. P. Dasgupta, P. David, Towards a new economics of science. *Res. Pol.* **23**, 486–521 (1994).
27. L. Shamir, The effect of conference proceedings on the scholarly communication in Computer Science and Engineering. *Scholarly Res. Commun.* **1**, 2 (2010).
28. M. Franceschet, The role of conference publications in computer science: A bibliometric view. *Comm. ACM* **53**, 129–132 (2010).
29. S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).
30. H. Sauerermann, C. Franzoni, Crowd science user contribution patterns and their implications. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 679–684 (2015).
31. P. Stephan, *How Economics Shapes Science* (Harvard Univ. Press, 2012).
32. K. Boudreau, Open platform strategies and innovation: Granting access vs. devolving control. *Manage. Sci.* **56**, 1849–1872 (2010).
33. R. S. Eisenberg, Proprietary rights and the norms of science in biotechnology research. *Yale Law J.* **97**, 177–231 (1987).
34. J. C. Flanagan, The critical incident technique. *Psychol. Bull.* **51**, 327–358 (1954).
35. L. D. Butterfield, W. A. Borgen, N. E. Amundson, A.-S. T. Maglio, Fifty years of the critical incident technique: 1954–2004 and beyond. *Qual. Res.* **85**, 475–497 (2005).

Acknowledgments: The institutional review boards (IRBs) of the Georgia Institute of Technology and the National Bureau of Economic Research both judged the research to present “no more than minimal risk of harm to subjects and to involve no procedures for which written consent is normally required outside of the research context.” The invitation email included a statement that completing the survey meant that the respondent had “read the information contained in [the explanatory] letter, and would like to be a volunteer in this research study.” All respondents to the survey were promised confidentiality. **Funding:** J.G.T. and M.C.T. acknowledge financial support from the NSF (award number 09652890). **Author contributions:** J.G.T. and C.H. contributed to all aspects of the research. J.G.T. conducted the sensitivity analysis and assembled the survey and publication data. C.H. analyzed the response bias. M.C.T. contributed to survey design, pretesting and analysis, and writing. L.J. contributed to survey design and writing. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data needed to evaluate the conclusions in the paper are presented in the paper and the Supplementary Materials. The IRB protocol for the survey required participant confidentiality and non-disclosure of individual-level data. Data from Scopus are available to researchers under API protocols as specified by Scopus. Interested readers should direct questions regarding the data to the corresponding author.

Submitted 29 October 2017

Accepted 3 April 2018

Published 16 May 2018

10.1126/sciadv.aar2133

Citation: J. G. Thursby, C. Haeussler, M. C. Thursby, L. Jiang, Prepublication disclosure of scientific results: Norms, competition, and commercial orientation. *Sci. Adv.* **4**, eaar2133 (2018).

Prepublication disclosure of scientific results: Norms, competition, and commercial orientation

Jerry G. Thursby, Carolin Haeussler, Marie C. Thursby and Lin Jiang

Sci Adv 4 (5), eaar2133.
DOI: 10.1126/sciadv.aar2133

ARTICLE TOOLS	http://advances.sciencemag.org/content/4/5/eaar2133
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2018/05/14/4.5.eaar2133.DC1
REFERENCES	This article cites 32 articles, 7 of which you can access for free http://advances.sciencemag.org/content/4/5/eaar2133#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the [Terms of Service](#)

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Advances* is a registered trademark of AAAS.