Research Policy 43 (2014) 1747–1759

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Scientific misbehavior in economics

Sarah Necker[a,b,*]

[a] University of Freiburg, Platz der alten Synagoge 1, 79098 Freiburg, Germany
[b] Walter-Eucken Institute, Goethestraße 10, 79100 Freiburg, Germany

ARTICLE INFO

Article history:
Received 15 August 2013
Received in revised form 16 April 2014
Accepted 14 May 2014
Available online 18 June 2014

Keywords:
Scientific misbehavior
"Publish or perish"
Social norms
Economics of crime
Academic labor market

ABSTRACT

This study reports the results of a survey of professional, mostly academic economists about their research norms and scientific misbehavior. Behavior such as data fabrication or plagiarism are (almost) unanimously rejected and admitted by less than 4% of participants. Research practices that are often considered "questionable," e.g., strategic behavior while analyzing results or in the publication process, are rejected by at least 60%. Despite their low justifiability, these behaviors are widespread. Ninety-four percent report having engaged in at least one unaccepted research practice. Surveyed economists perceive strong pressure to publish. The level of justifiability assigned to different misdemeanors does not increase with the perception of pressure. However, perceived pressure is found to be positively related to the admission of being involved in several unaccepted research practices. Although the results cannot prove causality, they are consistent with the notion that the "publish or perish" culture motivates researchers to violate research norms.

1. Introduction

The disclosure of scientific misbehavior usually causes great indignation by other scientists, the media, and the public. Even though scandals are rare, their existence fundamentally questions the image of science as a quest for truth. Scientific fraud distorts scholarly knowledge, thereby hampering and misleading scientific progress. However, trust in scientific research is also grounded on the assumption that it is unbiased by the researchers' presumptions or strategic behavior. In recent years, the focus on plagiarism and falsification has been given up in favor of an approach that also deals with questionable or "normal misbehavior" (De Vries et al., 2006). This includes practices applicable to a researcher's everyday tasks and goals. They can be as damaging to scientific progress as outright fraud. The aim of this paper is to provide evidence on fraudulent and questionable research practices in economics.

Any attempt to quantify the extent of scientific misbehavior must account for the issue that researchers have strong incentives to hide misbehavior. A popular approach is asking scientists directly. To the author's knowledge, four surveys gather evidence on certain types of misbehavior in economics (List et al., 2001; Enders and Hoover, 2004, 2006; Wilhite and Fong, 2012). The studies show that economics is subject to misbehavior. However, a broader knowledge of the research practices that economists consider unjustifiable and the prevalence of these practices is lacking.

An important question is why unaccepted research practices are employed. The most frequently cited cause is pressure to "publish or perish" (e.g., Frey, 2003; De Rond and Miller, 2005). Science has been compared to a winner-take-all market in which rewards are only granted to those first to make a discovery and therefore obtain recognition from peers (Stephan, 1996). The publication record has become increasingly important for survival in academia (e.g., Graber and Walde, 2008). At the same time, competition for publication space in top-journals has strongly increased (Franzoni et al., 2011; Card and DellaVigna, 2013). The economic theory of crime predicts that fierce competition increases the benefits of cheating (Becker, 1968). Fanelli (2010a) and Schwieren and Weichselbaumer (2010) provide evidence that cheating is more common in competitive environments. However, little is known about the link between pressure perceived by researchers and their misbehavior.

In autumn 2010, an anonymous online survey was conducted among the 2520 members of the European Economic Association (EEA), the professional body of economists from European and other countries.[1] In order to better understand which practices constitute scientific misbehavior, economists were asked to

* Correspondence to: University of Freiburg, Platz der alten Synagoge 1, 79098 Freiburg, Germany. Tel.: +49 761 790970.
E-mail address: sarah.necker@vwl.uni-freiburg.de

[1] A description of the results from a follow-up survey of German economists that had not participated in the survey of EEA members is provided in Necker (2012).

http://dx.doi.org/10.1016/j.respol.2014.05.002
0048-7333/

assess the justifiability of a wide range of behaviors. Respondents were surveyed about their own research practices and the expected and observed behavior of colleagues. The survey inquired into economists' perception of pressure and professional situation. A total of 631 economists responded to the survey; 426 continued until the last page.

This study examines the survey responses. Economists' acceptance of various research practices, the admitted use of those practices and evidence of misbehaving colleagues, as well as the perception of pressure are reported. The study provides the first examination of the link between perceived pressure and admitted misbehavior.

2. Previous literature

Extrapolating scientific misbehavior from disclosures is likely to capture only the tip of the iceberg (e.g., Steneck, 2006). The prevalence of positive or statistically significant results may be indicative of the use of questionable practices (e.g., Fanelli, 2010b; Brodeur et al., 2013). However, studies detecting irregularities only hint at anomalies and allow no inferences as to which practices were employed. A popular approach is the survey of researchers (e.g., Martinson et al., 2005; Titus et al., 2008).

The first survey of economists was conducted by List et al. (2001). The authors gathered information from the participants of the 1998 meetings of the American Economic Association. Falsification of data is admitted by 4.5% of the respondents. Up to 10% report having committed other misdemeanors such as having accepted or given unjustified co-authorship. The participants believe that 5–7% of research in the top 30 journals are falsified and that 13–17% are affected by other misdemeanors. List et al. (2001) cite the "Muhammad Ali effect" as an explanation for the discrepancy between admitted own behavior and the expected behavior of others. The term originates from the psychological literature. It is used to describe the finding that people see themselves as more moral than others (Allison et al., 1989).

The nature and extent of plagiarism are investigated by Enders and Hoover (2004). The authors surveyed editors from various economic journals. About 80% answer that unattributed sentences or the use of privately collected data without permission constitute plagiarism. Even more agree that this applies to unattributed proofs from working or published papers. In a typical year, a case of plagiarism is experienced by 29% of the responding editors. In a follow-up survey of "rank and file" economists almost one quarter reports that they have been plagiarized (Enders and Hoover, 2006). However, the authors stress that this number is likely to be biased by a disproportionately high participation rate of victims of plagiarism.

Wilhite and Fong (2012) analyze the prevalence of coercive citation. The authors surveyed researchers in economics, sociology, psychology, and multiple business disciplines in the US. Citation requests which are not based on a perceived omission in the academic content but on the desire to increase citations to the editor's home journal are considered inappropriate by 86% of the respondents. Having been asked to add citations for that reason is reported by 19.7%. Only 10% of those rejected the request. While the occurrence of coercion by journals in the fields of economics, psychology, and sociology is similar, journals in the business disciplines coerce more.

3. The survey

3.1. Survey methodology

When designing a survey on a topic like scientific misbehavior, it is important to minimize the likelihood that factors like perceived intrusiveness, fear of disclosure, or reluctance to admit socially undesirable attitudes or behavior affect the responses. Self-administration of questionnaires, in particular the privacy offered by an online survey, has been shown to decrease social desirability-biases (compared to personal interviews or paper forms) (Tourangeau and Yan, 2007). Confidentiality assurances have been found to increase unit and item response rates as well as the response quality when the requested data are sensitive (however, when the topic is non-sensitive assurances may arouse suspicion and backfire) (Singer et al., 1992, 1995).

The survey was administered online via SoSci Survey. The committee of the EEA sent an email to all members of the organization, inviting them to participate. The message included a non-personalized link to the questionnaire. Access was granted without a password. The invitation and introduction to the survey contained a confidentiality assurance in which anonymity was guaranteed. It was emphasized that the data would only be used in statistical analyses. The survey was accessible for eight weeks. Four weeks after the initial invitation a reminder was sent.

The first part of the survey dealt with the justifiability of different research practices. To get a broad picture of unacceptable behavior, various research practices were selected for evaluation. To explore the prevalence of misbehavior, respondents were asked whether they "have ever" engaged these practices and how many cases of researchers committing scientific misconduct they had direct evidence of. Finally, the perception of the research conditions and details on the personal situation were surveyed. The text of the questionnaire can be found in Section A of the Supplementary material (available online).

3.2. Non-response and representativeness

The population of the survey consists of all registered members of the EEA, i.e., 2520 individuals (assuming that all email addresses were available and correct). The questionnaire was started 631 times; 426 respondents continued until the last page. This amounts to a response rate of 25%, or 16.9% if dropouts are not counted. The sample is restricted to respondents who proceeded until the last page.

To check the representativeness of the sample, survey participants are compared to EEA members. Information on the location of workplace and gender are available for the sample and the population.[2] The sample is largely representative with respect to both characteristics. While 23% of survey participants are females, the EEA has 26.8% female members. The fraction of respondents working in a country corresponds to the fraction of members from that country, the correlation is 0.98.

An established procedure to check for biases caused by unit non-response is the comparison of responses from early and late respondents. The approach rests on the assumption that early respondents have a particular interest in the topic and answer the questionnaire immediately. Late respondents are assumed to be more similar to non-respondents since they would have fallen into that category if the reminder had not been sent (Armstrong and Overton, 1977). Two sample mean comparison tests show that the answers of the first and last quintile of respondents are similar.

The same result is obtained if responses from participants that continued until the last page and those that dropped out are compared. Two sample mean comparison tests show that with few

[2] The information on EEA members was kindly provided by Gemma Prunner-Thomas, EEA.

exceptions the justifiability of research practices is not significantly different between those that finished and those that dropped out.[3]

Two subsequent waves of the survey were conducted among members of the German and French economic associations.[4] In these waves, the questions on own (mis-)behavior were not posed to one sixth of randomly selected participants. The aim of this setup was to study whether surveying own behavior biases the answers. A comparison of responses from the randomly selected respondents to those from respondents answering all questions shows no significant differences (Necker, 2012).

Another issue is item non-response. Since item non-response rates are rather low (average missing rate 2.8%), all descriptive statistics are based on observed data. However, deleting observations with missing values implies an up to 14% smaller sample available for the empirical analysis. The missing values are filled in using an iterative multiple imputation (MI) procedure (Rubin, 1987, 1996). Five complete data sets were created. The details of the imputation are provided in Section B of the Supplementary material (available online).

4. Description of the data

4.1. Economists' norms

Respondents were asked "On a scale from 1 to 6 where 1 means "not at all justifiable (strongly agree)" and 6 means "highly justifiable (strongly disagree)," how do you assess the following behavior?" To test the internal consistency of the scale (interrelatedness of items), Cronbach's alpha is calculated. The α is 0.82. According to common rules of thumb, the value indicates high reliability of the scale.

Table 1 summarizes the responses. The survey reveals widespread agreement on how economists should choose a research topic. Intrinsic motivation is considered highly important; 85% (95% confidence interval (CI): 82–89%) agree that a topic should be chosen based on one's personal interest. Extrinsic motivation also plays an important role; 60% (CI: 55–64%) agree that the prospects for publication should be taken into account. With respect to topics for empirical research economists seem to follow a pragmatic approach; only 19% (CI: 16–23%) think that it is not acceptable to "define the research question according to data availability."

Economists clearly condemn behavior that misleads the scientific community or causes harm to careers. The least justifiable action is "copying work from others without citing." Respondents unanimously (CI: 99–100%) agree that this behavior is unjustifiable. Fabricating or correcting data as well as excluding part of the data are rejected by at least 97% (CI: 96–99%). "Using tricks to increase t-values, R^2, or other statistics" is rejected by 96% (CI: 94–98%), 93% (CI: 90–95%) consider "incorrectly giving a colleague co-authorship who has not worked on the paper" unjustifiable. Accordingly, some research norms are indeed fundamental and universal.

In the European Code of Conduct for Research Integrity (ESF and ALLEA, 2011) it is emphasized that there is a "thin borderline between some violations of [questionable research] practices and the serious types of misconduct." The survey shows that behavior typically listed as questionable are indeed only slightly more accepted. Not checking the contents of the works cited or not citing others' results if not in line with the analysis or from lower ranked journals is rejected by 86–91%. Practices such as searching for control variables until the desired result is found or selective presentation of those empirical findings that confirm one's argument are unaccepted by 81–85%. "Copying from own previous work without citing" is rejected by 80% (CI: 77–84%). Thus, it is much more accepted than plagiarism.

Strategic behavior in the publication process is also rejected but more accepted than practices applicable when analyzing data or writing papers. Citing strategically or maximizing the number of publications by slicing into the smallest publishable unit is rejected by 64% (CI: 60–69%). Complying with suggestions by referees even though one thinks they are wrong is considered unjustifiable by 61% (CI: 56–66%).

4.2. Economists' admitted own research practices

Participants were asked which research practices they had "ever" employed in the past. The responses are summarized in Table 2. Almost all economists (96%, CI: 94–97%) state that personal interest determined the choice of their research topic. The publication prospects were reported to have been decisive for 67% (CI: 62–71%).

The responses confirm the finding by List et al. (2001) that serious misbehavior exists in economics. The correction, fabrication, or partial exclusion of data, incorrect co-authorship, or copying of others' work is admitted by 1–3.5%. The use of "tricks to increase t-values, R^2, or other statistics" is reported by 7%. Having accepted or offered gifts in exchange for (co-)authorship, access to data, or promotion is admitted by 3%. Acceptance or offering of sex or money is reported by 1–2%. One percent admits to the simultaneous submission of manuscripts to journals.

About one fifth admits to having refrained from citing others' work that contradicted the own analysis or to having maximized the number of publications by slicing their work into the smallest publishable unit. Having at least once copied from their own previous work without citing is reported by 24% (CI: 20–28%). Even more admit to questionable practices of data analysis (32–38%), e.g., the "selective presentation of findings so that they confirm one's argument." Having complied with suggestions from referees despite having thought that they were wrong is reported by 39% (CI: 34–44%). Even 59% (CI: 55–64%) report that they have at least once cited strategically to increase the prospect of publishing their work.

According to their responses, 6.3% of the participants have never engaged in a practice rejected by at least a majority of peers. John et al. (2012) report almost the same fraction for psychologists. The authors find that 94% engaged in at least one questionable research practice (behavior similar to the ones considered here).

Table 2 also shows the average justifiability of research practices by respondents' own behavior. The results suggest a clear tendency of respondents who admit to a behavior to assign a higher level of justifiability to the research practice. The average difference is roughly one scale unit. John et al. (2012) report a similar difference.

4.3. Economists' reports of misbehaving colleagues

Respondents were asked which fraction of research in the top general and top field journals (A+or A) they believe to be subject to different types of misbehavior ("up to . . . %," scale given in deciles). The fabrication of data is expected to be the least widespread. The median response is "up to 10%." Respondents believe that incorrect handling of others' ideas, e.g., plagiarism, is more common; the median is "up to 20%" of published research. The "incorrect reporting of results, e.g., using tricks to improve statistics" as well

[3] Of those starting the survey, 17% left after visiting the first page (the introduction). Questions of "Part I: Norms" were answered by 56–38 dropouts.

[4] The aim was to reach economists that had not yet participated. Since follow-up surveys are only available for two countries, these data are disregarded in the analysis. Including the responses implies that about two thirds of the sample are either German or French. Results are largely unchanged when the observations are included, see Tables C.3 and C.4 of the Supplementary material (available online).

Table 1
Economists' norms.

No.	On a scale from 1 to 6. . .	Obs.	Ordinal variable				Binary variable			
			Mean	Std. dev.	[95% CI]		Mean	Std. dev.	[95% CI]	
General research approach: agreement										
1	A research topic should be chosen according to one's personal interest (in contrast to career concerns)	426	2.39	1.15	2.28	2.50	0.85	0.36	0.82	0.89
2	A research topic should be chosen with respect to publication prospect	426	3.33	1.20	3.21	3.44	0.60	0.49	0.55	0.64
3	A researcher should give credit to any published and unpublished idea by someone else (i.e., colleagues, journalists, students)	424	1.80	1.27	1.68	1.92	0.88	0.32	0.85	0.91
4	Results should be generalized if the theoretical framework or the research design for empirical analysis allows for it.	420	2.44	1.13	2.33	2.54	0.84	0.37	0.81	0.88
Research practices: justifiability										
5	Copying parts from the work of others without citing	426	1.06	0.30	1.03	1.09	1.00	0.05	0.99	1.00
6	Fabricating some data	417	1.19	0.74	1.12	1.26	0.97	0.17	0.96	0.99
7	Correcting data to fit the theory	423	1.23	0.65	1.17	1.30	0.98	0.14	0.96	0.99
8	Excluding part of the data (e.g., outliers) without reporting this	423	1.52	0.78	1.44	1.59	0.98	0.14	0.96	0.99
9	Using tricks to increase t-values, R^2, or other statistics	424	1.61	0.86	1.53	1.69	0.96	0.20	0.94	0.98
10	Incorrectly giving a colleague co-authorship who has not worked on the paper	425	1.83	1.04	1.73	1.93	0.93	0.26	0.90	0.95
11	Not citing results that are not in line with own analysis	426	2.01	1.17	1.90	2.12	0.89	0.31	0.86	0.92
12	Not checking the contents of the works cited	424	2.07	0.99	1.97	2.16	0.91	0.29	0.88	0.94
13	Not citing work in lower ranked journals, i.e., which in a ranking from A+to C rank lower than A	425	2.16	1.16	2.05	2.27	0.86	0.35	0.82	0.89
14	Presenting empirical findings selectively so that they confirm one's argument	424	2.19	1.17	2.08	2.30	0.84	0.37	0.81	0.88
15	Searching for control variables until you get the desired results	422	2.21	1.18	2.10	2.32	0.85	0.36	0.82	0.88
16	Stopping statistical analysis when you have a desired result	423	2.45	1.24	2.33	2.56	0.81	0.40	0.77	0.84
17	Copying from your own previous work without citing	425	2.47	1.33	2.34	2.59	0.80	0.40	0.77	0.84
18	Not citing work from other disciplines	421	2.58	1.27	2.46	2.70	0.77	0.42	0.73	0.81
19	Citing strategically to raise publication prospects (e.g., to please editors or possible referees)	426	3.00	1.31	2.87	3.12	0.64	0.48	0.60	0.69
20	Maximizing the number of publications by dividing the work into the smallest publishable units, meaning several individual articles covering similar topics and differing from each other only slightly	426	3.06	1.26	2.94	3.18	0.64	0.48	0.60	0.69
21	Complying with suggestions from referees or editors when you think they are wrong	424	3.09	1.39	2.96	3.23	0.61	0.49	0.56	0.66
22	Defining the research question according to data availability	422	4.51	1.23	4.39	4.63	0.19	0.40	0.16	0.23

The scale was given as agreement (strongly agree: 1 – strongly disagree: 6) with respect to the general research approach (items 1–4). For research practices (items 5–22) it was given as justifiability (not at all justifiable: 1 – highly justifiable: 6). Items 5–22 are listed in increasing order of justifiability. Binary variable is set to unity if respondent chose responses 1–3 of the rating scale. Based on observed data.

as "incorrect application of empirical methods, e.g., data mining" are assumed to be more prevalent. The median is "up to 30%."

Having observed at least one case of "scientific misconduct" in the department or institute is reported by 146 persons (34%, CI: 29–38%). About half of those observed only one case, two cases were observed by 18%, three cases by 6% and 22.6% report having observed more than three cases. To decrease the likelihood of duplicate reporting, respondents were asked about the cases they had observed "in their department or institute." To increase the comparability across researchers, the restriction "in the past 3 years" was given. The question did not define "scientific misconduct." Affirmative responses are based on the respondent's assessment.

Respondents that answered affirmatively were asked about the details of their observation. With regard to the type of misconduct, 18% (based on all respondents) report that the problem they had observed was unsound handling of others' ideas. Fourteen percent report that they had observed unsound application of empirical

Table 2
Economists' admitted own research behavior.

No.	Have you ever . . .	Obs.	Percent "Yes"	Std. dev.	[95% CI]		Norms by behavior		Δ	Difference significant?
							Not admitted	Admitted		
General research approach										
1	Chosen a research topic according to your personal interests	423	95.51	20.74	93.53	97.49	3.32	2.35	−0.97	***
2	Chosen a research topic with respect to publication prospect	423	66.90	47.11	62.40	71.41	3.92	3.05	−0.87	***
3	Built your research on someone else's idea without giving credit	420	1.90	13.69	0.59	3.22	1.80	1.63	−0.17	–
4	Generalized your results even though the theoretical framework or the research design for empirical analysis did not allow for it	419	13.13	33.81	9.88	16.37	2.42	2.49	0.07	–
Research practices										
5	Copied parts from work of others without citing	422	2.13	14.46	0.75	3.52	1.05	1.67	−0.62	***
6	Fabricated some data	348	2.59	15.90	0.91	4.26	1.14	3.22	−2.09	***
7	Corrected data to fit the theory	348	1.15	10.67	0.02	2.27	1.21	1.21	−0.79	***
8	Excluded part of the data (e.g., outliers) without reporting this	348	3.45	18.27	1.52	5.37	1.47	3.08	−1.62	***
9	Used tricks to increase t-value, R^2, or other statistics	348	7.18	25.86	4.46	9.91	1.52	2.64	−1.12	***
10	Failed to correctly give a colleague co-authorship who has worked on the paper	423	1.42	11.84	0.29	2.55	1.82	1.50	0.32	–
11	Refrained from citing results or opinions that are not in line with your own analysis	422	21.09	40.84	17.18	25.00	1.83	2.67	−0.85	***
12	Refrained from checking the contents of the works cited	422	51.90	50.02	47.11	56.68	1.62	2.48	−0.86	***
13	Refrained from citing work in lower ranked journals, which in a ranking from A+to C rank lower than A	421	19.95	40.01	16.12	23.79	1.94	2.94	−1.00	***
14	Presented empirical findings selectively so that they confirm one's argument	348	32.18	46.79	27.25	37.12	1.82	2.99	−1.17	***
15	Searched for control variables until you got the desired results	348	36.49	48.21	31.41	41.58	1.75	2.94	−1.20	***
16	Stopped statistical analysis when you had a desired result	348	37.93	48.59	32.81	43.05	1.94	3.23	−1.28	***
17	Copied from your own previous work without citing	423	23.64	42.54	19.58	27.71	2.18	3.37	−1.20	***
18	Refrained from citing work from other disciplines	419	19.57	39.72	15.76	23.38	2.40	3.24	−0.84	***
19	Cited strategically to raise publication prospects (e.g., to please editors or possible referees)	420	59.29	49.19	54.57	64.00	2.35	3.43	−1.08	***
20	Maximized the number of publications by dividing the work to the smallest publishable unit, meaning several individual articles covering similar topics and differing from each other only slightly	423	19.86	39.94	16.04	23.68	2.84	3.89	−1.05	***
21	Complied with suggestions by referees or editors when you thought that they are wrong	420	39.05	48.84	34.36	43.73	2.67	3.74	−1.07	***
22	Defined the research question according to data availability	345	79.13	40.70	74.82	83.44	3.90	4.72	−0.82	***
23	Submitted a manuscript simultaneously to two or more journals violating journal policies	424	1.18	10.81	0.15	2.21	–	–	–	–
24	Accepted or offered gifts in exchange for (co-)authorship, access to data or promotion of particular persons	425	3.29	17.87	1.59	4.99	–	–	–	–
25	Accepted or offered money in exchange for (co-)authorship, access to data or promotion of particular persons	424	1.41	11.82	0.28	2.54	–	–	–	–
26	Accepted or offered sex in exchange for (co-)authorship, access to data or promotion of particular persons	425	1.88	13.61	0.58	3.17	–	–	–	–

The research practices are listed in increasing order of justifiability. The number of observations is roughly 17% lower with respect to items 6–9, 14–16, and 22 as those respondents indicated that they do not work empirically. The justifiability of practices 23–26 was not inquired. Some institutions, e.g., journals or grant giving institutions, have introduced respective codes of conduct. Based on observed data.

∗ Significance level: 10%

∗∗ Significance level: 5%.

∗∗∗ Significance level: 1%.

methods in their environment. Twelve percent accuse colleagues of having inappropriately presented results. Four percent report having witnessed the fabrication of data. Five percent report that the problem was incorrect co-authorship.

Fifty percent (based on affirmative responses) report that the researcher suspected of scientific misbehavior was a full professor. Compared to their fraction in the sample (27%), full professors are disproportionally often accused of misconduct. The involvement of PhD students is reported by 22%, of assistant/junior professors by 20%, of associate professors by 12%, and of post docs or researchers by 10%.

Twenty-four percent respond that they reported the suspected misconduct. Notification to the governing body of the institution is reported by 31% (based on those that made allegations). Having shared the observation with the media is reported by 25%. Twenty percent state that they took action by talking informally to other scholars about the issue. Notification to an ethics committee is reported by one respondent – supporting the view that statistics published by those institutions capture only the tip of the iceberg.

The suspicions of misbehavior are reported to have proved accurate in 62.5% of allegations. Participants report that instances of misconduct led to consequences most often when the governing body of the institution was informed (10 in 11 cases). Only 5 of the 14 cases in which colleagues were informed resulted in consequences.

Respondents indicating that they did not report the observed misconduct were asked for their reason. Thirty percent of those that refrained from allegations state that they did not know to whom they should report. Consequences for themselves were feared by 24%. For 18% the reason was loyalty towards the offender.

Respondents were asked more specifically whether they had observed certain types of misbehavior. A complaint of having been incorrectly excluded as co-author is voiced by 18%. The same fraction reports having observed others simultaneously submitting a manuscript to more than one journal, thus violating publication rules. A remarkable finding is that the majority of researchers that report having observed these behaviors respond negatively when asked whether they had observed "misconduct." While economists strongly agree on the importance of giving due credit to the contribution of others, the results may be taken as an indication that some tolerate the exploitation of their own ideas. However, as described in Section 2, the survey by Enders and Hoover (2004) shows that editors' agreement that an action constitutes the serious offense of plagiarism depends on the material that is unattributed. For example, unattributed sentences are considered less serious. The incorrect exclusion as co-author may be considered not serious enough to be classified as "misconduct."

4.4. Economists' perception of pressure

Economists were surveyed regarding their perception of the pressure to publish and to raise external funds. The respondents almost unanimously agree that publication pressure exists. Publication pressure is perceived to be "very high" by 38.9%, and "high" by 44.3%, as shown in the upper panel of Fig. 1. The lower panel of the figure shows that the perceived pressure to raise external funds is lower but still substantial. It is "very high" for 11%, "high" for 35.6%, and "moderately high" for 36.5%. Ten percent state that they do not perceive pressure to raise external funds.

Economists strongly agree that pressure has increased over the last decade, as shown in Fig. 2. A fraction of 45.5% (46.8%) believes that publication pressure (pressure to raise external funds) has "increased"; 33.7% (21.5%) even perceive a "strong increase." Constant pressure to publish (raise external funds) is reported by 7.2% (9.6%). Hardly any respondent states that pressure has decreased.

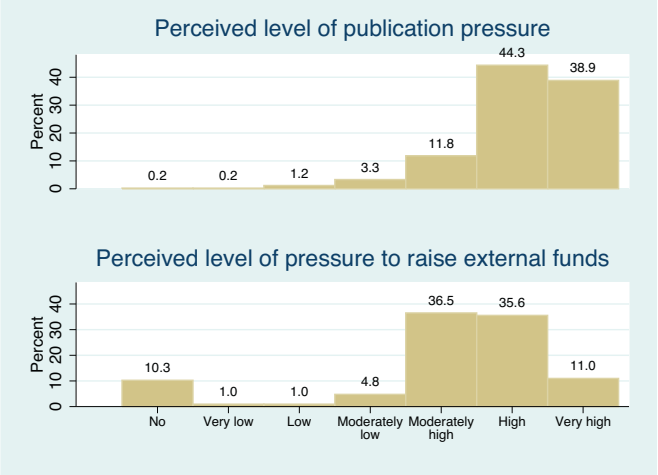

Fig. 1. Perceived level of pressure. *Note*: Based on 422/419 observations.

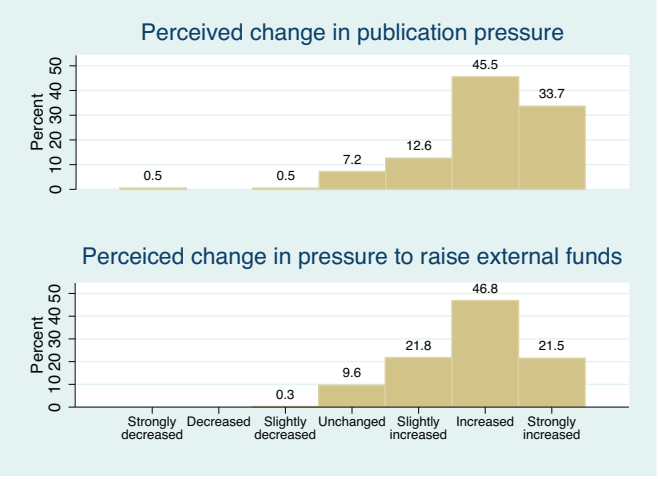

Fig. 2. Perceived change of pressure. *Note*: Based on 405/406 observations.

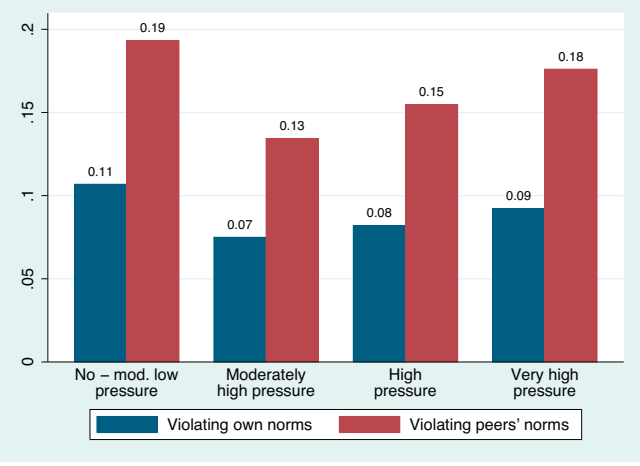

Fig. 3. Violations of research norms by perceived publication pressure. *Note*: Based on 422 observations. Violation of own norms = number of admitted misbehavior/number of rejected behavior; misbehavior classified as such if economist himself rejects the practice (responses 1–3 on a scale from 1 = not at all justifiable to 6 = highly justifiable). Violation of others' norms = same calculation; misbehavior classified as such if majority of peers reject the practice (at least 50% chose responses 1–3).

Fig. 3 plots perceived publication pressure against two variables indicating the respondent's admitted misbehavior. The first variable captures how many behavior the economist reports to have engaged in on the behavior he rejects, i.e., the violations of the economist's own convictions. The second variable captures an economist's violations of peers' research norms, i.e., engagement in practices rejected by at least 50% of participants. The figure indicates that the fraction of violated research norms is always lower if based on an economist's own convictions. The highest fraction of violations of own and others' research norms is reported by economists perceiving "no – moderately low" pressure. However, only a small fraction reports such low pressure. With regard to economists perceiving higher pressure, a positive relationship between pressure and violating research norms is indicated. The fraction of research norms (own and peers') that is reported to have been violated is slightly higher among respondents that perceive stronger pressure. With regard to peers' norms, the fractions are statistically different from each other.

5. Empirical analysis

5.1. Empirical approach

The relationship between economists' perception of pressure and scientific misbehavior is studied in an empirical analysis. Perceptions might be related to the reported research practices as well as the norms that researchers hold. In the one case, the dependent variable is the justifiability the respondent assigns to a behavior (ordinal variable, 1 = "not at all justifiable," . . ., 6 = "highly justifiable"). In the other case, admitted own research behavior is used as the dependent variable (binary variable, set to unity if the respondent reports having engaged in behavior, zero otherwise).

The models are estimated for the justifiability and own use of all research practices that are rejected by a majority (rejection rates shown in Table 1). However, low variance of the dependent variables prevents convergence of the models estimated for practices 5–10 or causes perfect prediction of several observations. These practices are summarized in one variable "(justifiability of) serious misbehavior."

The main explanatory variable is the respondent's perception of publication pressure. Being evaluated based on the scientific performance and having a tenured position are included as indirect measures of publication pressure. The economic theory of crime assumes that the decision to cheat is influenced by the expected probability of detection (Becker, 1968). The expectation of others' behavior has been found to also determine the likelihood of cheating (List et al., 2001) The regressions control for the expected probability of detection and the expected prevalence of misbehavior.

To explain the decision to cheat, economists have paid increasing attention to norms and internal rewards (e.g., Elster, 1989). Merton (1979) stresses that researchers internalize norms of science and act on them. Wilhite and Fong (2012) and John et al. (2012) report a positive link between the acceptance of a research practice and its admission. The respondent's stated view on the justifiability of a behavior is included in regressions on admitted own behavior. It is of course difficult to establish causality. Rather than behaving according to norms, people might equally well have adjusted their values to their behavior (Akerlof and Dickens, 1982).

All regressions control for the respondent's gender, year of birth, the academic position, country of workplace, and the fraction of time a researcher indicates as available for doing research. The ordinal and nominal controls are included as a set of dummies. The reference group is the highest or lowest category in case of ordered alternatives, among these the more frequent category is chosen. The reference group of unordered alternatives is the mode. Descriptions and summary statistics of all control variables are provided in Table C.1 of the online Supplementary material.

The five multiply imputed complete data sets are used separately for the analysis, the results are combined using Rubin's rule, i.e., such that they reflect the uncertainty due to the imputation.[5] The determinants cannot be said to be truly exogenous. It is impossible to make claims about the direction of causality.

5.2. Results from regressions explaining economists' norms

Table 3 provides results from ordered probit regressions on the justifiability the respondent assigns to the different unaccepted research practices. Shown are average marginal effects on the outcome "not at all justifiable." The results suggest that perceived pressure is negatively related to the justifiability of some misbehavior. The reference group is respondents perceiving "very high" pressure. Respondents perceiving less publication pressure assign higher justifiability to several research practices. For example, the probability that a respondent assesses "copying one's owns work" to be "not at all justifiable" is 2 ppts lower if he perceives "high" pressure, 7 ppts lower if he perceives "moderately high" pressure and 14 ppts lower if he perceives "no – moderately low" pressure. Most effects are statistically insignificant. Nonetheless, the result contradicts expectations.[6] However, the main hypothesis is that perceived pressure causes dishonest behavior, not a dishonest attitude.

Table 3 shows that being evaluated based on the scientific performance, i.e., having actual pressure to publish, increases the acceptance of some research practices. For example, the probability that a respondent strongly rejects "searching for control variables until the desired result is found" is 16.9 ppts lower if he is evaluated based on the scientific performance. Research norms are unrelated to having tenure.

Economists' normative convictions are significantly related to gender and year of birth (age and cohort cannot be distinguished). The probability that men find the analyzed behavior "not at all justifiable" is on average 5.7 ppts lower (results not reported, available upon request). Men have been found to respond more strongly to the competitiveness of their environment (e.g., Croson and Gneezy, 2009) and to be more likely to cheat (e.g., Friesen and Gangadharan, 2012). Fig. 4 depicts the marginal effects of year of birth in box plots. It shows that, compared to those born after 1979, economists born earlier are increasingly likely to find the listed behavior unjustifiable. Little evidence is found for cross-country differences in economists' norms (compared to Germany, i.e., the largest group, results available upon request).

5.3. Results from regressions explaining admitted research behavior

Table 4 shows average marginal effects from probit regressions on admitted own research behavior. The results suggest that the admission of several research practices is positively related to

[5] The MI estimate of β is $\overline{\beta}_M = (1/5)\sum_{i=1}^{5}\hat{\beta}_i$. The variance–covariance matrix of $\overline{\beta}_M$ is $T = \overline{U} + (1 + (1/5))B$ where $\overline{U} = \sum_{i=1}^{5}(\hat{U}_i/5)$ is the within-imputation variance–covariance matrix and $B = \sum_{i=1}^{5}((\beta_i - \overline{\beta}_M)(\beta_i - \overline{\beta}_M)')/(5-1)$ is the between-imputation variance–covariance matrix. Results are largely unchanged if only observed data are used. Available upon request.

[6] To check the robustness of the finding, the set of dummies capturing perceived publication pressure is replaced by different variables, e.g., an ordinal or binary variable capturing pressure. The results point to the same conclusion as the reported ones. The modifications are described in detail in Section 5.3 with respect to admitted behavior. Results available upon request.

Table 3
Regressions on economists' norms.

Description	Dependent variable					
No. of behavior	Serious misbehavior 5–10 AME/SE	Disregard results 11 AME/SE	No check of content 12 AME/SE	Cite low ranked 13 AME/SE	Present selectively 14 AME/SE	Search controls 15 AME/SE
Perceived publication pressure...						
...no – moderately low	−0.062	−0.075	−0.231***	−0.053	−0.090	−0.204***
	(0.045)	(0.100)	(0.058)	(0.089)	(0.078)	(0.062)
...moderately high	−0.013	0.020	−0.086	−0.008	0.009	−0.104*
	(0.013)	(0.068)	(0.059)	(0.070)	(0.063)	(0.057)
...high	−0.010	0.069	0.023	0.008	0.012	0.024
	(0.007)	(0.046)	(0.047)	(0.046)	(0.042)	(0.047)
...very high	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Has tenured position	−0.003	−0.053	0.037	0.036	0.060	−0.012
	(0.011)	(0.074)	(0.066)	(0.066)	(0.056)	(0.063)
Scientific performance is evaluated	−0.016	−0.000	−0.050	−0.097*	−0.078	−0.169***
	(0.010)	(0.061)	(0.048)	(0.056)	(0.052)	(0.050)
Perceived probability of detection...						
...very low	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
...low	0.003	0.070	−0.013	0.035	0.042	0.030
	(0.013)	(0.065)	(0.076)	(0.075)	(0.063)	(0.072)
...moderately low	−0.006	0.052	0.015	0.014	0.021	−0.026
	(0.014)	(0.063)	(0.074)	(0.071)	(0.065)	(0.068)
...high	0.006	−0.011	0.032	0.133	0.125	0.082
	(0.014)	(0.085)	(0.090)	(0.088)	(0.084)	(0.084)
Expects high prevalence of misbehavior	−0.009	−0.009	0.027	0.047	−0.033	−0.001
	(0.007)	(0.043)	(0.041)	(0.040)	(0.037)	(0.039)
Other controls?	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	67.08	40.93	70.76	50.85	78.06	47.00
Pseudo R^2	0.04	0.03	0.05	0.04	0.06	0.04
N	426	426	426	426	426	426

Description	Dependent variable					
No. of behavior	Stop analysis 16 AME/SE	Copy own work 17 AME/SE	Cite oth. disc. 18 AME/SE	Cite strategic. 19 AME/SE	Max. publications 20 AME/SE	Comply suggestions 21 AME/SE
Perceived publication pressure...						
...no – moderately low	−0.141**	−0.142**	−0.054	−0.039	0.010	−0.040
	(0.057)	(0.064)	(0.061)	(0.045)	(0.046)	(0.047)
...moderately high	−0.014	−0.066	0.064	0.002	0.029	−0.017
	(0.052)	(0.051)	(0.057)	(0.036)	(0.032)	(0.037)
...high	0.056	−0.021	0.021	0.008	0.032	−0.012
	(0.038)	(0.040)	(0.036)	(0.027)	(0.022)	(0.024)
...very high	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Has tenured position	−0.028	0.009	0.008	−0.027	0.034	0.016
	(0.055)	(0.053)	(0.054)	(0.042)	(0.033)	(0.036)
Scientific performance is evaluated	−0.079*	0.042	0.015	−0.037	−0.001	−0.067**
	(0.046)	(0.051)	(0.044)	(0.031)	(0.029)	(0.031)
Perceived probability of detection...						
...very low	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
...low	0.021	0.068	−0.018	0.052	0.015	0.041
	(0.059)	(0.059)	(0.053)	(0.032)	(0.031)	(0.036)
...moderately low	−0.018	0.006	0.046	0.073**	0.045	0.025
	(0.057)	(0.058)	(0.053)	(0.032)	(0.032)	(0.034)
...high	0.084	0.066	0.059	0.151***	0.060	0.092*
	(0.073)	(0.074)	(0.068)	(0.053)	(0.042)	(0.050)
Expects high prevalence of misbehavior	−0.016	−0.075**	0.074**	0.012	−0.030	−0.021
	(0.034)	(0.035)	(0.033)	(0.024)	(0.020)	(0.023)
Other controls?	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	51.12	43.38	46.14	67.25	71.79	70.34
Pseudo R^2	0.04	0.03	0.03	0.05	0.05	0.05
N	426	426	426	426	426	426

Ordered probit estimates, reported are average marginal effects (AME) on the outcome "not at all justifiable." The variable "justifiability of serious misbehavior" assumes the value of 0 if all of the practices 5–10 are considered "not at all justifiable," the value of 1 if one of these behaviors is considered justifiable, and so on. All 5 imputations are used, results combined using Rubin's rule. Hypothesis tests based on robust standard errors. Measures of fit are the lowest statistic among results from the five imputations. (ref.), category is reference group. Other controls: location of workplace, male, year of birth, rank, fraction of research time.

* Significance level: 10%.
** Significance level: 5%.
*** Significance level: 1%.

Table 4
Regressions on admitted research behavior.

Description	Dependent variable					
No. of behavior	Serious misbehavior 5–10 AME/SE	Disregard results 11 AME/SE	No check of content 12 AME/SE	Cite low ranked 13 AME/SE	Present selectively 14 AME/SE	Search controls 15 AME/SE
Perceived publication pressure...						
...no – moderately low	0.017	−0.034	−0.029	0.071	−0.124	−0.080
	(0.075)	(0.089)	(0.100)	(0.100)	(0.084)	(0.100)
...moderately high	−0.056	−0.083	−0.169***	−0.127**	−0.142**	−0.083
	(0.045)	(0.054)	(0.061)	(0.052)	(0.061)	(0.076)
...high	0.035	0.012	−0.033	−0.023	−0.110**	−0.028
	(0.040)	(0.043)	(0.048)	(0.041)	(0.050)	(0.049)
...very high	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Has tenured position	0.042	−0.077	0.046	0.020	0.010	0.060
	(0.052)	(0.059)	(0.068)	(0.055)	(0.075)	(0.066)
Scientific performance is evaluated	0.024	−0.083	0.079	0.087	−0.010	0.028
	(0.051)	(0.052)	(0.061)	(0.061)	(0.055)	(0.059)
Perceived probability of detection...						
...very low	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
...low	0.049	0.011	0.012	0.048	−0.045	−0.108
	(0.050)	(0.070)	(0.079)	(0.053)	(0.069)	(0.085)
...moderately low	0.062	−0.047	0.031	0.096*	0.013	−0.109
	(0.054)	(0.070)	(0.080)	(0.053)	(0.070)	(0.084)
...high	−0.009	−0.145**	−0.004	0.012	−0.105	−0.078
	(0.055)	(0.074)	(0.089)	(0.064)	(0.082)	(0.101)
Expects high prevalence of misbehavior	0.124***	−0.025	−0.068	0.036	0.005	0.058
	(0.036)	(0.040)	(0.045)	(0.039)	(0.044)	(0.043)
Justifiability of practice (1 = not at all, 6 = high)	0.047***	0.092***	0.233***	0.106***	0.146***	0.183***
	(0.011)	(0.014)	(0.019)	(0.014)	(0.017)	(0.017)
Other controls?	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	56.11	60.10	121.97	77.54	101.97	92.83
Pseudo R^2	0.16	0.15	0.23	0.18	0.30	0.26
N	387	426	426	426	353	352

Description	Dependent variable					
No. of behavior	Stop analysis 16 AME/SE	Copy own work 17 AME/SE	Cite oth. disc. 18 AME/SE	Cite strategic. 19 AME/SE	Max. publications 20 AME/SE	Comply suggestions 21 AME/SE
Perceived publication pressure...						
...no – moderately low	−0.050	0.000	−0.024	−0.125	−0.076	−0.029
	(0.119)	(0.091)	(0.084)	(0.105)	(0.079)	(0.116)
...moderately high	−0.188***	−0.048	−0.044	−0.130*	0.023	−0.036
	(0.060)	(0.054)	(0.057)	(0.069)	(0.060)	(0.067)
...high	−0.062	−0.008	0.018	−0.108**	0.006	0.021
	(0.048)	(0.041)	(0.043)	(0.047)	(0.038)	(0.048)
...very high	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Has tenured position	−0.040	0.072	0.058	−0.034	−0.004	−0.034
	(0.067)	(0.054)	(0.057)	(0.070)	(0.056)	(0.065)
Scientific performance is evaluated	−0.084	0.010	0.024	0.001	0.144***	0.079
	(0.059)	(0.048)	(0.051)	(0.057)	(0.054)	(0.065)
Perceived probability of detection...						
...very low	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
...low	0.009	0.041	0.090	−0.094	−0.033	−0.188**
	(0.078)	(0.061)	(0.061)	(0.074)	(0.066)	(0.083)
...moderately low	0.054	0.028	0.072	−0.138*	−0.003	−0.251***
	(0.082)	(0.061)	(0.061)	(0.074)	(0.065)	(0.083)
...high	−0.159*	0.025	−0.033	−0.184**	−0.002	−0.286***
	(0.087)	(0.072)	(0.062)	(0.091)	(0.078)	(0.095)
Expects high prevalence of misbehavior	−0.039	0.103***	0.009	0.024	0.090**	0.101**
	(0.044)	(0.037)	(0.039)	(0.045)	(0.036)	(0.044)
Justifiability of practice (1 = not at all, 6 = high)	0.172***	0.103***	0.078***	0.135***	0.105***	0.117***
	(0.015)	(0.012)	(0.014)	(0.015)	(0.014)	(0.014)
Other controls?	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	117.48	115.14	69.45	118.93	86.80	94.27
Pseudo R^2	0.32	0.24	0.15	0.25	0.22	0.18
N	353	426	426	426	426	426

Probit estimates, reported are average marginal effects (AME). The variable "serious misbehavior" is set to unity if the respondent employed at least one serious misbehavior (practices 5–10), and zero otherwise. All 5 imputations are used, results combined using Rubin's rule. Hypothesis tests based on robust standard errors. Measures of fit are the lowest statistic among results from the five imputations. (ref.), category is reference group. Other controls: location of workplace, male, year of birth, rank, fraction of research time. Number of observations lower for behavior where respondent could indicate "I do not work empirically."

* Significance level: 10%.
** Significance level: 5%.
*** Significance level: 1%.

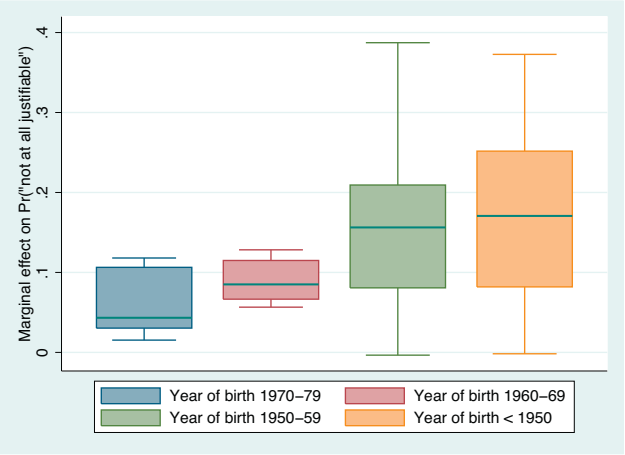

Fig. 4. Year of birth and rejection of misbehavior. *Note*: Each box plot based on 12 average marginal effects from the 12 regressions shown in Table 3. Average marginal effects on the outcome "not at all justifiable." Reference group is respondents born after 1979.

perceived publication pressure. Compared to respondents perceiving "very high" pressure, the perception of only "high" pressure implies, e.g., an 11 ppts lower probability of admitting to have "selectively presented results" or "cited strategically." The probability that respondents perceiving only "moderately high" pressure admit some behavior is even lower. For example, reporting to have "selectively presented results" is 14 ppts less likely, the effect is 17 ppts with respect to "not having checked a work's contents" and 19 ppts with respect to having "stopped the analysis when the desired result was found" (significant at 1/5% level).

The effects of lower perceived publication pressure are largely negative and of similar magnitude also with regard to other behavior. While these effects are only partially statistically significant, it has to be kept in mind that most respondents perceive some degree of high pressure (see Fig. 1). Few observations in a category, e.g., respondents perceiving "no" or "low" pressure, decrease the statistical power of the result. Results that are based on few observations are unlikely to be statistically significant. The sample sizes corresponding to each combination of perceived publication pressure and behavior are shown in Table C.2 of the Supplementary material (available online).

To check the robustness of the finding, first, the set of dummies capturing perceived publication pressure is replaced by the corresponding ordinal variable. Second, the set of dummies is replaced by a binary variable which is set to unity if the respondent perceives "very high" pressure, zero otherwise. The results, shown in Table 5, indicate again a positive association between perceived pressure and admitted misbehavior. The pattern of results resembles the one shown in Table 4.

To analyze whether the change in publication pressure over the past decade is related in a similar way to admitted misbehavior, the perception of the change is included instead of the perceived level. As shown in Table 5, the perception of a higher increase in pressure is positively related to admitting misbehavior. The effects are again only partially statistically significant. However, it has to be considered that economists agree largely also about the change of pressure (see Fig. 2). In line with previous results, a statistically significant link is found, e.g., with respect to the practices of early stopping of the empirical analysis and citing strategically.

Furthermore, the perceived pressure to raise external funds is included instead of the pressure to publish. The results, shown in Table 5, indicate a positive link also between perceived pressure to raise external funds and admitted misbehavior. However, the effects differ across practices to a larger extent and indicate a different pattern. For example, a relationship to admitting serious misbehavior is indicated which is not the case for publication pressure. Admitting strategic citation is positively related to publication pressure but not to the pressure to raise external funds.

Table 4 shows that the justifiability a respondent assigns to a behavior is clearly positively related to his admitted own behavior. A one standard deviation-increase in justifiability increases the likelihood that a respondent reports not having checked contents, having searched for control variables, or having stopped the analysis when the desired result was found by 21–23 ppts. An effect of 16–18 ppts is found with respect to the selective presentation of results, the compliance with unconvincing referee suggestions, or strategic citation. Admitting to not having cited lower-ranked journals, having maximized the number of publications, or having copied one's own work is 12–13 ppts more likely. The effect is 10–11 ppts with respect to not having cited other disciplines or results not in line with one's own analysis. The lowest relationship is found with respect to serious misbehavior (7 ppts).

In line with the results by List et al. (2001), the expectation of a high prevalence of misbehavior is significantly positively related to the admission of several practices. Results on respondents' location of workplace (not reported, available upon request) suggest that admission to misbehavior differs to some extent across countries.

6. Discussion

The survey contained several provisions to increase the accuracy of responses. Compared to other surveys of economists on the same topic, the generated response rate (17%) is in the middle of the range. While the paper-based surveys by List et al. (2001) and Enders and Hoover (2004) yielded a rate of 23 and 28% of usable responses, the web-based surveys by Enders and Hoover (2006) and Wilhite and Fong (2012) yielded a lower return (7 and 9–12% of usable responses, respectively). The response rates of web surveys have been found to be usually lower than those of other survey modes (Kaplowitz et al., 2004; Manfreda et al., 2008). However, as emphasized by Krosnick (1999), "it is not necessarily true that representativeness increases monotonically with increasing response rate. [. . .] surveys with very low response rates can be more accurate than surveys with much higher response rates."

The examination of the representativeness of the sample and the responses does not indicate systematic biases. However, this does not prove that unit non-response, self-deception, or dishonest answers do not affect the results. For questions that allow a direct comparison, self-reports are much lower than observations of others' behavior (affirmative answers: 1% versus 18%). To some extent, this may be the consequence of duplicate reporting of cases. The fraction of reported misbehavior is usually found to be much larger when it is asked about others' than about the researcher's own behavior (e.g., John et al., 2012). However, the difference may also be caused by dishonest responses or the "Muhammad Ali effect." Fanelli (2009) argues that "while surveys asking about colleagues are hard to interpret conclusively, self-reports systematically underestimate the real frequency of scientific misconduct."

In their study of questionable research practices in psychology, John et al. (2012) provide a truth-telling incentive (donation to a charity depending on the truthfulness of responses) to two-thirds of respondents. Respondents without those incentives less often admit some questionable research practices. The difference to respondents with incentives is high for falsifying data, moderate for practices such as premature stopping of data collection, and negligible for six other, more accepted practices. It would be plausible if biases were larger with respect to less accepted actions.

Respondents were asked whether they have "ever" engaged in the listed activity. Frequency of engagement was not inquired

Table 5
Regressions on admitted research behavior – modifications.

Description	Dependent variable					
	Serious misbehavior	Disregard results	No check of content	Cite low ranked	Present selectively	Search controls
No. of behavior	5–10	11	12	13	14	15
	AME/SE	AME/SE	AME/SE	AME/SE	AME/SE	AME/SE
I. Perceived publication pressure as ordinal variable (no-very high)						
Ordinal publication pressure	0.006	0.021	0.042*	0.018	0.061**	0.033
	(0.020)	(0.024)	(0.025)	(0.023)	(0.025)	(0.028)
Other controls as in Table 4	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	52.01	57.92	113.13	66.76	100.20	92.73
Pseudo R^2	0.15	0.15	0.22	0.17	0.29	0.26
N	387	426	426	426	353	352
II. Perceived publication pressure binary: "very high" vs. less pressure						
Perceived publ. pressure "very high"	−0.016	0.010	0.057	0.035	0.116**	0.044
	(0.037)	(0.039)	(0.044)	(0.038)	(0.046)	(0.047)
Other controls as in Table 4	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	52.00	57.92	113.13	75.86	100.20	92.73
Pseudo R^2	0.15	0.15	0.22	0.17	0.29	0.26
N	387	426	426	426	353	352
III. Perceived change of publication pressure. . .						
. . .unchanged/decreased	−0.141***	−0.097	0.050	0.038	−0.120	−0.232***
	(0.047)	(0.066)	(0.083)	(0.082)	(0.105)	(0.088)
. . .slightly increased	−0.020	−0.002	−0.041	−0.086	−0.125*	−0.040
	(0.060)	(0.067)	(0.069)	(0.056)	(0.065)	(0.078)
. . .increased	−0.073*	−0.002	−0.013	−0.061	−0.050	−0.070
	(0.040)	(0.047)	(0.050)	(0.045)	(0.050)	(0.051)
. . .strongly increased	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Other controls as in Table 4	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	56.11	60.10	121.97	73.74	101.97	92.83
Pseudo R^2	0.16	0.15	0.23	0.18	0.29	0.26
N	382	425	425	425	352	351
IV. Perceived pressure to raise external funds. . .						
. . .no – moderately low	−0.169**	−0.001	0.048	0.007	−0.074	−0.009
	(0.077)	(0.082)	(0.086)	(0.071)	(0.091)	(0.089)
. . .moderately high	−0.123	−0.030	−0.017	0.016	−0.096	−0.017
	(0.075)	(0.075)	(0.079)	(0.065)	(0.076)	(0.076)
. . .high	−0.175**	−0.060	0.038	0.055	−0.056	−0.014
	(0.072)	(0.073)	(0.077)	(0.063)	(0.078)	(0.076)
. . .very high	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Other controls as in Table 4	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	62.83	59.89	111.19	67.75	101.53	95.84
Pseudo R^2	0.18	0.15	0.22	0.17	0.28	0.26
N	387	426	426	426	352	352

Description	Dependent variable					
	Stop analysis	Copy own work	Cite oth. disc.	Cite strategic.	Max. publications	Comply suggestions
No. of behavior	16	17	18	19	20	21
	AME/SE	AME/SE	AME/SE	AME/SE	AME/SE	AME/SE
I. Perceived publication pressure as ordinal variable (no-very high)						
Ordinal publication pressure	0.060**	0.011	0.011	0.058**	0.009	0.010
	(0.027)	(0.023)	(0.023)	(0.026)	(0.023)	(0.028)
Other controls as in Table 4	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	113.86	112.35	66.91	114.90	85.94	90.73
Pseudo R^2	0.31	0.24	0.15	0.25	0.22	0.18
N	353	426	426	426	426	426
II. Perceived publication pressure binary: "very high" vs. less pressure						
Perceived publ. pressure "very high"	0.085*	0.016	−0.004	0.113**	−0.001	−0.006
	(0.045)	(0.038)	(0.040)	(0.044)	(0.036)	(0.045)
Other controls as in Table 4	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	113.86	112.35	47.71	114.90	36.03	90.73
Pseudo R^2	0.31	0.24	0.15	0.25	0.09	0.18
N	353	426	426	426	426	426
III. Perceived change of publication pressure						
. . .unchanged/decreased	−0.075	−0.116*	−0.104	−0.072	0.117	−0.060
	(0.110)	(0.067)	(0.075)	(0.089)	(0.081)	(0.086)
. . .slightly increased	−0.153**	−0.023	−0.100*	−0.210***	−0.019	−0.051
	(0.069)	(0.061)	(0.060)	(0.075)	(0.058)	(0.069)
. . .increased	−0.139***	−0.041	−0.045	−0.100**	−0.033	0.039
	(0.050)	(0.044)	(0.046)	(0.047)	(0.041)	(0.052)

Table 5 (*Continued*)

Description	Dependent variable					
	Stop analysis	Copy own work	Cite oth. disc.	Cite strategic.	Max. publications	Comply suggestions
No. of behavior	16	17	18	19	20	21
	AME/SE	AME/SE	AME/SE	AME/SE	AME/SE	AME/SE
. . .strongly increased	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Other controls as in Table 4	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	117.48	113.26	69.15	118.93	86.80	94.27
Pseudo R^2	0.32	0.24	0.15	0.25	0.22	0.18
N	352	425	425	425	425	425
IV. Perceived pressure to raise external funds. . .						
. . .no – moderately low	−0.159**	0.003	−0.159**	−0.026	0.000	−0.194**
	(0.077)	(0.070)	(0.080)	(0.087)	(0.068)	(0.087)
. . .moderately high	−0.033	0.018	−0.093	0.008	−0.052	−0.077
	(0.067)	(0.062)	(0.072)	(0.079)	(0.059)	(0.081)
. . .high	−0.021	−0.002	−0.150**	0.095	−0.036	−0.125
	(0.067)	(0.062)	(0.070)	(0.079)	(0.056)	(0.080)
. . .very high	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Other controls as in Table 4	Yes	Yes	Yes	Yes	Yes	Yes
Chi^2	95.84	114.47	75.20	108.22	87.58	94.52
Pseudo R^2	0.31	0.22	0.17	0.22	0.23	0.19
N	352	426	426	426	426	426

Panels I: include ordinal pressure variable instead of set of dummies, higher value = higher pressure. Panels II: include binary variable that is set to unity if respondent perceives "very high" publication pressure instead of set of dummies. Panels III: include perceived change of publication pressure instead of level. Panels IV: include perceived pressure to raise external funds instead of pressure to publish. Other explanations: see notes to Table 4.

about. John et al. (2012) report that when frequency response scales are used, 64% of affirmative responses fall into the "once/twice" category, 26% fall into "occasionally," and 10% report employing practices "frequently." The authors find that frequency scales more often lead to affirmative answers than dichotomous ones. This provides another reason why self-reports are a conservative estimate of the prevalence of misbehavior.

Regarding others' behavior it was broadly asked whether respondents had observed "scientific misconduct." Respondents that did not classify colleagues' behavior as "misconduct" did not answer subsequent questions. The responses suggest that economists had serious misbehavior in mind when responding affirmatively. From that perspective, others' misbehavior may also be underreported.

To get an idea of whether economists report misbehavior more often than other researchers, the results are compared to studies from other disciplines. It has to be taken into account that the variance between studies may partly be due to different modes of delivery, objects of study (self or others), and wording (Fanelli, 2009).

Self-reports and reports on others' behavior indicate that, relative to other practices, falsification and fabrication are a rare phenomenon in economics (1–4%). At least with respect to self-reports, high consistency seems to exist across studies (List et al., 2001; John et al., 2012; Fanelli, 2009). However, the meta-analysis by Fanelli (2009), covering surveys from different disciplines, shows that on average 14.1% accuse colleagues. Scientists from the management disciplines "observed or heard about" 27% of colleagues employing data falsification (Bedeian et al., 2010).

Having engaged in questionable research practices is reported by 20–59% of economists. The prevalence of comparable practices is similar in psychology (36–78%, John et al., 2012). Between 50 and 91% of management scholars report knowledge of colleagues employing questionable research practices (Bedeian et al., 2010). The surveyed economists relatively often admit strategic behavior in the publication process (39–59%). Only 25% of management scholars report that they made revisions to conform to editors' or referees' preferences (Bedeian, 2003). Having given in to coercive citation is reported by 18% of social scientists (Wilhite and Fong, 2012).

An important question is what should and can be done to tackle scientific misbehavior in economics. Even though institutions have been established that handle allegations of scientific misbehavior, the survey reveals that half of the respondents that report having observed misconduct either did not know to whom to report or feared consequences for themselves. While complainants may not directly accrue rewards, they face costs and the potential consequences of accusing a colleague. Deficiencies in the protection of whistle blowers are, e.g., reported by Reich (2011). Hoover (2006) shows in a model that due to the low chance of success, opportunity and legal costs, fighting is unlikely to be worthwhile for the original author of a plagiarized article. As emphasized also by Hoover (2006), increasing the awareness of independent bodies that hear complaints about misbehavior and the protection for whistle blowers seem to be important steps to fight scientific misconduct.

With respect to questionable research practices, it has to be considered that the actions do not necessarily represent intentional bias. Researchers have to make several decisions while analyzing data or writing down results. Stopping the analysis when the desired result has been found may be the consequence of ambiguity about the most reliable model. Some decisions are unforeseeable and impossible to make beforehand. "Not citing work from other disciplines" may be due to carelessness rather than deliberate neglect. However, malleable categorization has been found to facilitate the re-interpretation of behavior in a self-serving manner (Mazar et al., 2008). The difference in the strength of the link between justifiability and admitted behavior found in the analysis across practices may be explained by the possibility to categorize an action in terms that are compatible with good conduct.

Glaeser (2006) proposes that "researcher initiative bias" should be accepted and results be adjusted accordingly. The strong rejection of several research practices voiced by the survey participants casts doubts on whether this is a desirable strategy. The simulations by Simmons et al. (2011) show how easily "researcher degrees of freedom" can translate into statistically significant evidence for a false hypothesis. Some research practices may be used due to a lack of awareness that they are highly rejected. Mazar et al. (2008) and Pruckner and Sausgruber (2013) show experimentally that moral reminders increase honest behavior. The prevalence of questionable practices may be limited by reminding researchers, e.g., in

submission procedures, that engaging in those practices represents scientific misbehavior.

7. Conclusion

To the author's knowledge, this is the first study providing evidence for the justifiability and prevalence of a variety of research practices in economics. The survey reveals a broad consensus among professional economists about the norms that should guide researchers' behavior. The extent to which economists admit having employed unaccepted behavior is noteworthy. As in any other study of this topic, it has to be considered that the survey requested sensible information. The prevalence of misbehavior is likely to be biased – probably downward. The most frequently cited cause for scientific misbehavior is publication pressure. The surveyed economists strongly agree that high publication pressure exists. The study provides the first examination of the relationship between perceived pressure and admitting misbehavior. The results indicate that the perception of pressure is positively related to the admission of several research practices rejected by a majority of economists.

Acknowledgements

The author is grateful to the editors and three anonymous referees for excellent comments and suggestions. The author is indebted to Lars P. Feld and Bruno S. Frey. Without their support, the project would have been impossible. Thanks to Tim Besley and Gemma Prunner-Thomas for their support and the possibility to conduct the survey among EEA members. Heiko Burret, Axel Dreher, Jürgen Hoffmeyer-Zlotnik, Lars Kaczmirek, Dominic Leiner, Benoit Le Maux, Natalja Menold, and Reinhold Necker gave valuable comments. The author benefited also from comments of participants of the faculty seminar at the University of Rennes 1, Beyond Basic Questions Workshop, Congress of the EEA, Congress of the Verein für Socialpolitik, and Empirical Economics Seminar at the University of Freiburg (all of them in 2012).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.respol.2014.05.002.

References

Akerlof, G.A., Dickens, W.T., 1982. The economic consequences of cognitive dissonance. American Economic Review 72 (3), 307–319.
Allison, S.T., Messick, D.M., Goethals, G.R., 1989. On being better but not smarter than others: the Muhammad Ali effect. Social Cognition 7 (3), 275–295.
Armstrong, J., Overton, T., 1977. Estimating nonresponse bias in mail surveys. Journal of Marketing Research 14, 396–402.
Becker, G.S., 1968. Crime and punishment: an economic approach. Journal of Political Economy 76 (2), 169–217.
Bedeian, A.G., 2003. The manuscript review process – the proper roles of authors, referees, and editors. Journal of Management Inquiry 12 (4), 331–338.
Bedeian, A.G., Taylor, S.G., Miller, A.N., 2010. Management science on the credibility bubble: cardinal sins und various misdemeanors. Academy of Management Learning and Education 9 (4), 715–725.
Brodeur, A., Lé, M., Sangnier, M., Zylberberg, Y., 2013. Star Wars: The Empirics Strike Back. IZA Working Paper No. 7268.
Card, D., DellaVigna, S., 2013. Nine Facts About Top Journals in Economics. National Bureau of Economic Research Working Paper No. 18665.
Croson, R., Gneezy, U., 2009. Gender differences in preferences. Journal of Economic Literature 47 (2), 448–474.
De Rond, M., Miller, A.N., 2005. Publish or perish – bane or boon of academic life? Journal of Management Inquiry 14 (4), 321–329.
De Vries, R., Anderson, M.S., Martinson, B.C., 2006. Normal misbehavior: scientists talk about the ethics of research. Journal of Empirical Research on Human Research Ethics 1 (1), 43–50.
Elster, J., 1989. Social norms and economic theory. Journal of Economic Perspectives 3 (4), 99–117.
Enders, W., Hoover, G., 2006. Plagiarism in the economics profession: a survey. Challenge 49 (5), 92–107.
Enders, W., Hoover, G.A., 2004. Whose line is it? Plagiarism in economics. Journal of Economic Literature 42 (2), 487–493.
ESF, ALLEA, 2011. European Code of Conduct for Research Integrity. IREG, Strasbourg.
Fanelli, D., 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS ONE 4 (5), e5738.
Fanelli, D., 2010a. Do pressures to publish increase scientists' bias? An empirical support from US states data. PLoS ONE 5 (4), e10271.
Fanelli, D., 2010b. "Positive" results increase down the hierarchy of the sciences. PLoS ONE 5 (4), e10068.
Franzoni, C., Scellato, G., Stephan, P., 2011. Changing incentives to publish. Science 333, 702–703.
Frey, B.S., 2003. Publishing as prostitution? Choosing between one's own ideas and academic success. Public Choice 116 (1–2), 205–223.
Friesen, L., Gangadharan, L., 2012. Individual level evidence of dishonesty and the gender effect. Economics Letters 117 (3), 624–626.
Glaeser, E.L., 2006. Researcher Incentives and Empirical Methods. National Bureau of Economic Research Technical Working Paper No. 329.
Graber, M., Walde, A.L., 2008. Publish or perish? The increasing importance of publications for prospective professors in Austria, Germany and Switzerland. German Economic Review 9, 457–472.
Hoover, G.A., 2006. A game-theoretic model of plagiarism. Atlantic Economic Journal 34 (4), 449–454.
John, L.K., Loewenstein, G., Prelec, D., 2012. Measuring the prevalence of questionable research practices with incentives. Psychological Science 23 (5), 524–532.
Kaplowitz, M.D., Hadlock, T.D., Levine, R., 2004. A comparison of web and mail survey response rates. Public Opinion Quarterly 68 (1), 94–101.
Krosnick, J.A., 1999. Survey research. Annual Review of Psychology 50 (1), 537–567.
List, J.A., Bailey, C.D., Euzent, P.J., Martin, T.L., 2001. Academic economists behaving badly? A survey on three areas of unethical behavior. Economic Inquiry 39 (1), 162–170.
Manfreda, K., Bosnjak, M., Berzelak, J., Haas, V.V.I., 2008. Web surveys versus other survey modes. International Journal of Market Research 50 (1), 79–104.
Martinson, B.C., Anderson, M.S., De Vries, R., 2005. Scientists behaving badly. Nature 435 (7043), 737–738.
Mazar, N., Amir, O., Ariely, D., 2008. The dishonesty of honest people: a theory of self-concept maintenance. Journal of Marketing Research 45 (6), 633–644.
Merton, R.K., 1979. The Sociology of Science: Theoretical and Empirical Investigations. University of Chicago Press.
Necker, S., 2012. Wissenschaftliches Fehlverhalten – Ein Problem in der deutschen Volkswirtschaftslehre? Perspektiven der Wirtschaftspolitik 13, 267–285.
Pruckner, G., Sausgruber, R., 2013. Honesty on the streets: a field study on newspaper purchasing. Journal of the European Economic Association 11 (3), 661–679.
Reich, E.S., 2011. Whistle-blower claims his accusations cost him his job. Nature 474, 140–141.
Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. Wiley, New York.
Rubin, D.B., 1996. Multiple imputation after 18+years. Journal of the American Statistical Association 91 (434), 473–489.
Schwieren, C., Weichselbaumer, D., 2010. Does competition enhance performance or cheating? A laboratory experiment. Journal of Economic Psychology 31 (3), 241–253.
Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science 22 (11), 1359–1366.
Singer, E., Hippler, H.-J., Schwarz, N., 1992. Confidentiality assurances in surveys: reassurance or threat? International Journal of Public Opinion Research 4 (3), 256–268.
Singer, E., Von Thurn, D.R., Miller, E.R., 1995. Confidentiality assurances and response: a quantitative review of the experimental literature. Public Opinion Quarterly 59 (1), 66–77.
Steneck, N.H., 2006. Fostering integrity in research: definitions, current knowledge, and future directions. Science and Engineering Ethics 12 (1), 53–74.
Stephan, P.E., 1996. The economics of science. Journal of Economic Literature 34 (3), 1199–1235.
Titus, S.L., Wells, J.A., Rhoades, L.J., 2008. Repairing research integrity. Nature 453 (7198), 980–982.
Tourangeau, R., Yan, T., 2007. Sensitive questions in surveys. Psychological Bulletin 133 (5), 859–883.
Wilhite, A.W., Fong, E.A., 2012. Coercive citation in academic publishing. Science 335 (6068), 542–543.