Pathogenic organization in science: Division of labor and retractions

John P. Walsh[a],[*], You-Na Lee[b], Li Tang[c]

[a] *Georgia Institute of Technology, Atlanta, Georgia, United States*
[b] *National University of Singapore, Singapore*
[c] *Fudan University, China*

ARTICLE INFO

ABSTRACT

Science is increasingly a team activity, and the size of the teams has been growing. At the same time, there are concerns about an increasing rate of pathologies in science. The growth of team science suggests the need to look beyond individual-level explanations and focus on organizational structures and institutional contexts to explain pathologies in science. Drawing on the literature on organizational pathologies, we argue that division of labor may be a key factor contributing to pathologies in science. Furthermore, we examine the effects of high-stakes incentives and of institutional corruption as additional predictors of scientific pathologies. Using retractions as an indicator of pathologies, and drawing on a matched sample of 195 retracted papers and 349 paired papers that were not retracted, we develop indicators of the division of labor in the team that produced a paper and find that the rate of retractions is higher as the division of labor increases (net of team size). Additionally, we find that high-stakes incentives and institutional corruption are also associated with increased retractions. We conclude with a discussion of the implications of these findings for science policy, in particular for organizing team science projects.

The comparative student of man's work learns about doctors by studying plumbers; and about prostitutes by studying psychiatrists—Everett C. Hughes, "Mistakes at Work"

…[O]ne can begin to glimpse the sources, other than idiosynchratic ones, of the misbehavior of individual scientists. The culture of science is, in this measure, pathogenic.–Robert K. Merton, "Priorities in Scientific Discovery"

1. Introduction

While research pathologies (falsification, fabrication, significant error, etc.) in science are a longstanding problem (Merton, 1973), there are concerns that the rates of such pathologies have been increasing in recent years. (Azoulay et al., 2015b; Baron et al., 2016; Harley et al., 2014; Honig et al., 2014). At the same time, recent decades have also produced an increase in team sizes in science (Milojević, 2014; Wuchty et al., 2007), which suggests that the study of science may benefit from incorporating an organization theory perspective. In this paper, we develop an organization theory of scientific pathologies that links the structure of collaborations with the likelihood of pathologies. Prior work on organizational pathologies suggests that pathologies are not just driven by individual failings or cultural environment ("bad apples

or bad barrels", cf. Brass et al. (1998)), but rather are also endemic to organizations. Furthermore, this literature suggests that the rates of pathologies vary systematically with organizational structure, in particular with the degree of division of labor (Greve et al., 2010; Vaughan, 1999). Based on this prior work, we develop an organization theory of scientific pathologies. We also examine institutional contexts, including incentive systems and institutional corruption in a society. We will then test the implications of this theory using data on retractions in biomedical and life sciences publications to demonstrate empirical relations consistent with our theory. We find that a higher degree of division of labor is associated with greater likelihood of retraction, consistent with our organization theory of scientific pathologies.

We motivate our discussion with some illustrations. In 1986 Nobel Laureate David Baltimore, Thereza Imanishi-Kari and others published a paper in *Cell* with the dramatic finding that transplanted genes can change a host animal's anti-body production. Soon after, a post-doctoral researcher, Margot O'Toole, at Tufts, investigated the original data and reported irregularities in the Imanishi-Kari lab notebooks. The result was a series of investigations into possible fraud, eventually including high profile Congressional hearings. The paper was retracted in 1991. In addition, Baltimore resigned his position as president of Rockefeller University. In his response to the Office of Scientific Integrity report on the case (Baltimore, 1991), Baltimore claimed that, although he was a

---

co-author and advisor on the paper, he did not closely follow the data collection or analysis, and could not know if it was fraudulent or not: "I wish to state that if Imanishi-Kari did falsify data or make mis-representations, I had no knowledge of the misconduct." Baltimore was supervising the first author, Weaver, who was Baltimore's post-doc, and who used Imanishi-Kari's data. As Baltimore described it, "The study that gave rise to the paper was conducted as a classic collaboration, with each laboratory performing independent research in its particular area. … Imanishi-Kari provided the expertise in serology that I lacked." As we will show below, such division of labor and specialization may make scientific projects vulnerable to mistakes or to intentional mis-representation. Eventually the case was declared a result of error and Baltimore and Imanishi-Kari were cleared of accusations of wrongdoing (for a detailed discussion of this case, see Chubin and Hackett (1990)).

A more recent high-profile case of pathology in science is the re-traction of a 2014 *Science* paper on the impact of door-to-door can-vassing for changing people's fundamental beliefs (Carey and Belluck, 2015). The first author, Michael LaCour, then a political science doc-toral student at UCLA, has admitted to data fabrication and falsifica-tion. His senior co-author, Donald P. Green, a professor at Columbia University, agreed to help guide LaCour. Green was initially skeptical and asked LaCour to replicate the findings, which he did with a follow-up study. Presented with these results, Green focused on statistical analyses, helping with writing the results and contributing his ability to craft the findings for higher impact. He did not, however, examine the original data or participate in the survey design or data collection: "Convinced that the results were robust, I helped Michael LaCour write up the findings, especially the parts that had to do with the statistical interpretation of the experimental design. Given that I did not have IRB approval for the study from my home institution, I took care not to analyze any primary data — the datafiles that I analyzed were the same replication datasets that Michael LaCour posted to his website. Looking back, the failure to verify the original Qualtrics data was a serious mistake." (Carey and Belluck, 2015).

Another example, from solid state physics, is the case of Jan Hendrik Schön at Bell Labs. Schön was accused of fabrication and falsification on a large number of papers and an internal committee at Bell Labs found that he had committed scientific misconduct (Beasley et al., 2002). Schön defended himself against many of the allegations, arguing in-stead that these were honest mistakes, or attempts to present cleaner graphs to better illustrate the phenomena he observed, or simply sloppy record-keeping. Most of the papers that were investigated were co-authored. Generally these co-authors provided materials that were used as part of Schön's experiments. However, the committee stated clearly that none of the co-authors had committed misconduct, although there were some questions about whether they had failed in their professional responsibilities. As the report stated, "Except for the provision of starting materials by others, all device fabrication, physical measure-ment and data processing in the work in question were carried out (with minor exceptions) by Hendrik Schön alone, with no participation by any coauthor or other colleague. None of the most significant phy-sical results was witnessed by any coauthor or other colleague. (pp2-3)" (Beasley et al., 2002). Regarding the responsibilities of co-authors, the committee accepted that there is often division of labor and specialized expertise in a collaboration and rejected the position that all co-authors are responsible for all aspects of the publication. They did note that "coauthors often have access to technical details that other parties, such as management, referees, editors and award committees do not have, and thus the coauthors represent the first line of defense against mis-conduct. When that defense fails, as in this case, it emphatically raises the question of whether the community has a right to expect more from coauthors. (p. 16)" (Beasley et al., 2002). The report goes on to discuss the high level of trust that is assumed among co-authors, and the dif-ficulties that explicitly doubting and confronting colleagues would create for maintaining this necessary trust. This example illustrates both the vulnerability of a project with a clear division of labor (in this case

between those providing samples and the one designing the studies, doing the experiments, analyzing the data and generating the figures that are the basis for the write-up), and the need for co-authors to take responsibility to ensure that the results they are publishing on are technically sound.

Then there is the notorious case of Diederik Stapel, an eminent Dutch social psychologist who was accused of falsifying or fabricating data on over 50 publications (Bhattacharjee, 2013). Stapel was able to continue with his deceptions for many years, in part because he did not let his co-authors collect data or do the analyses, not even letting his graduate students do the analyses although he had risen to the rank of dean.

In contrast to this extreme case of clear misbehavior by an eminent scientist, Paté-Cornell presents a hypothetical example of a civil en-gineer taking a soil measurement incorrectly, and this getting passed on to the design engineer who develops the building plans based on this faulty information, leading later to the oil rig's collapse (Paté-Cornell, 1990).

A key point in these several examples is that, in some of these cases (the Stapel case, for example), there is a clear culprit. In others, it is not clear that anyone involved misbehaved (save for "honest mistakes"). And, in some cases, there was significant disagreement, even after substantial investigation, about whether it was one or the other (the Baltimore case). However, in each case, "the rig collapsed": the validity of the project's results turns out to be in doubt. Hence, we use the term scientific pathologies to describe these cases where the validity of the scientific finding turns out to be in doubt because of problems with the project, without recourse to the underlying individual behaviors that might be blamed for the problem (after the fact). We discuss this de-finitional issue in more detail below.

These examples suggest that division of labor might in fact increase the vulnerability of projects to pathologies, since other researchers are unlikely to know the details of how that part of the study was done. While cultural and institutional factors, not to mention individual proclivities, are clearly important (cf. Lewellyn et al., 2017), we will argue that the organization of the project provides important insights into the likelihood of pathologies (cf. Baron et al., 2016; Lewellyn et al., 2017).

In the following sections, we discuss pathologies in science, in-cluding definitions of pathologies and prior work on predictors of such pathologies. We then develop an organization theory of scientific pathologies that provides novel insights into pathologies in science beyond commonly used individual and cultural explanations. We then show empirical evidence consistent with the implications of this theory, using retracted and non-retracted papers in biomedical and life sciences to measure pathologies and the contribution statements published in those papers to measure division of labor. The main finding shows that a greater division of labor is associated with a higher likelihood of re-traction, net of several other factors associated with retractions (in-cluding incentives, levels of societal corruption and interdisciplinarity). Finally, we conclude with a discussion of the theoretical and policy implications.

2. Pathologies in science

2.1. Defining pathologies in science

Those studying pathologies in science use a wide variety of defini-tions and terms (covering overlapping sets of behaviors), including: misconduct, error, fraud, deviance, research malpractice, selective re-porting, p-hacking, almost wrongs, false science, detrimental research practices, etc. (Azoulay et al., 2015b; Chubin, 1985; Chubin and Hackett, 1990). The Office of Research Integrity [ORI] uses the fol-lowing definition of "research misconduct": "fabrication, falsification or plagiarism in proposing, performing or reviewing research, or in re-porting research results" (42 C.F.R. Part 93). ORI further states that

"….Research misconduct does not include honest error or differences of opinion." Much of this work emphasizes an absolutist standard, and, often, the need for assessing blame; and emphasizes a psychological or legalistic definition of "misconduct", often revolving around "intent" (National Academy of Sciences, 2017). The National Academy of Sciences [NAS] also points out that operationalizing such definitions can be difficult. This skepticism about the ability to clearly demarcate "misconduct" from "questionable research practice" from "honest mistake", or similar categorizations is echoed in reviews of the study of misconduct in the sociology of science (Butler et al., 2017; Hackett, 1994; Zuckerman, 1988). Zuckerman's (1988) discussion of "reputable" and "disreputable" errors highlights the difficulties of the absolutist view, since even without intent, "honest mistakes", when seen by social control agents as being outside the bounds of the cognitive norms of acceptable research practices, can lead to a loss of credibility and to sanctions (such as retraction). The lengthy review process and reversals in the Baltimore and Imanishi-Kari case illustrate this difficulty (Chubin and Hackett, 1990).

Instead of using an absolutist definition based on a universal moral code and establishment of intent, we adopt a sociological definition of deviance, building on labeling theory (Becker, 1963) and organizational misconduct research (cf. Greve et al., 2010; Warren, 2003). Warren (2003) notes the tension in the management literature between those who focus on specific deviant behaviors (Merton, 1949) and those who focus on acts that become labeled as deviant (Becker, 1963). Furthermore, the management literature on deviance in organizations emphasizes both a negative meaning (misconduct) and a positive meaning (innovation). We argue that the focus should be on the pathological outcomes without the necessity of assuming intent, or limiting to specific sub-categories of behaviors (i.e., beyond ORI's fabrication, falsification, plagiarism categorization). Even the National Academies' recent report on fostering research integrity notes the need to expand beyond these ORI categories, adding the concept of "detrimental research practices" (which includes significant errors and lack of supervision) to their discussion of research pathologies (NAS, 2017).

Employee deviance can be associated with both desirable and undesirable behaviors, and whether a particular behavior is desirable or undesirable may depend on who is making that judgment (Warren, 2003). For example, while we argue below that division of labor is associated with higher rates of pathologies, Lee et al. (2015) show that division of labor is also associated with higher scientific novelty. Warren argues for a reference group perspective, noting that there are local and external reference groups and specific behaviors can be congruent or incongruent with the norms or expectations of one or the other (cf. Mohliver, 2018). Warren also notes that these norms sometimes conflict, so that we can ask if certain norms are themselves problematic. Hence, pathologies can sometimes be the result of engaging in behaviors that are consistent with local norms but contradict external norms. Mohliver (2018) argues that such liminal behaviors spread through local networks. Similarly, we might expect that liminal standards for conducting scientific research (that may be susceptible to being redefined as pathological) are also likely to spread through invisible colleges (Crane, 1972). There are many examples in the world of business of such liminal behaviors that are locally defined as acceptable but later declared deviant by social control agents (see Warren, 2003; Greve et al., 2010; Mohliver, 2018).

Therefore, Greve et al. (2010) argue that rather than defining deviance as particular behaviors (often theorized as originating in particular individual or cultural characteristics), it is more fruitful to define deviance based on labeling theory (Becker, 1963). As stated by Greve et al. (2010): "We define organizational misconduct as behavior in or by an organization that a social-control agent judges to transgress a line separating right from wrong; where such a line can separate legal, ethical, and socially responsible behavior from their antitheses. (p. 56)". Thus, a particular behavior is pathological not because it is inherently antithetical to a normative or moral principle, but rather

because it has caught the attention and the approbation of an agent of social control. From this definition, the same behavior can in one case be accepted and in another be defined as deviant, because of the actions of social control agents. For example, Mohliver's (2018) study of stock-option backdating highlights that there was significant ambiguity about the appropriateness of this liminal practice and that it was initially viewed favorably by many local networks of auditors, to the point where it became standard practice in some teams in the industry, and only later becomes redefined by social control agents as illegitimate. We can see similar cases in the world of science (Chubin and Hackett, 1990). For example, discarding data points or methods of analysis as unreliable outliers may be viewed as supportive of the lab's research agenda to produce high quality findings (and hence be viewed as consistent with the lab's goals and norms). However, this same behavior may be viewed as negative behaviors worthy of punishment by other reference groups. The debate regarding the Baltimore paper is a classic example. One advantage of this labeling theory definition is that it is operationally tractable. We know that particular behaviors are deviant because they have been labeled as such (Becker, 1963). Over time and in other circumstances, the same behavior, even if known, might be tolerated or even lauded (Mohliver, 2018). In the context of science, particular research practices only become deviant when they are labeled "misconduct" by, for example, ORI; or lead to a retraction by a journal (both being agents of social control). Chubin (1985) uses a similar labeling theory perspective to understand pathologies in science. Both Warren's perspective and Greve's (and Chubin's) perspective reject an absolutist view of research pathologies. And, both perspectives have merit, depending on the research question at hand. For the study of the organizational covariates of pathologies, we favor Greve's interpretation. As Greve, et al. note, this definition has the advantage of being empirically verifiable: researchers have engaged in pathologies when they have been declared so by the journal (or some other social control agent). Greve et al. explicitly include state actors (such as ORI) and professional bodies (such as peer judgments by journal editors) as social control agents. Hence, we will define pathologies as retracted papers, meaning a journal has nullified the paper. In other words, the focus is on the outcome (the nullified paper), and not on adjudicating blame for that outcome, i.e. without recourse to (often contested) judgments about whether or not this falls into one of the ORI categories, or whether this was intentional malfeasance or unfortunate accident, or even whether this judgment was the result of initiative by one of the authors or a third party. Greve and others (e.g., Perrow, 1984; Vaughan, 1999) note that pathologies need not involve intent (in contrast to ORI's definition), but that the complexities and interdependencies in an organization can produce bad outcomes even in cases where all involved are faithfully performing their roles (we discuss this further below). Hence, we will not distinguish between different motivations (or labelings) of the "types" of retractions. Rather, our theories will point to structural causes of both intentional and unintentional factors that result in pathologies. In addition, we are focusing on outcomes that affect scientific results (the results have been nullified, although they may still turn out to be true), what Robinson and Bennett term "organizational harms", rather than inter-personal harms (Robinson and Bennett, 1995). As Robinson and Bennet point out, it is important to clarify the scope of deviant outcomes that are being considered in order to reduce the heterogeneity of cases and likely causes. Hence, we are not focusing on what the NAS (2017) refers to as "other misconduct" (for example, sexual harassment, misuse of funds, etc.).

In part our choice of a labeling theory of pathologies in science is made based on the purpose of the analysis. In our theory of team structure and pathologies, we make similar predictions for various forms of pathologies (intentional and unintentional). On the other hand, for those who study the *effects* of pathologies of different types, a definition that distinguishes intentional malfeasance from accidental error may be more useful (cf. Azoulay, et al., 2015a,b). Greve et al. (2010) do note that one limitation of our definition is that we cannot

readily distinguish whether the covariates predicting pathologies are predicting the likelihood of engaging in particular behaviors or the likelihood of being examined and labeled by social control agents (or some combination). Hence, work that could make this distinction would help move this area of research forward. For now, we focus on the factors that contribute to papers being labeled as involving pathologies, broadly defined (meaning, all those behaviors that can result in a paper being retracted).

2.2. Rates and effects of pathologies in science

Because pathologies in science may harm the fundamental structure of science as a cumulative process and the public's trust in, and support for, science, there is significant policy focus on such pathologies, especially as there are concerns that such pathologies are widespread and growing (NAS, 2017). For example, Fig. 1 shows that from 1985 to 2010 the rate of retractions for papers published in PubMed goes from about 0.4 per 10,000 publications to about 4.0 per 10,000, a 10-fold increase in 25 years. Similarly, in 2013 the NSF Inspector General reported that over the last 10 years the office observed a tripling of allegations received about, and findings of, misconduct per year (Lerner, 2013).

Most studies use widely available measures of pathologies, such as retractions (as we will do here). A few studies attempt to "audit" pathologies using plagiarism detection software or similar comparisons (Honig and Bedi, 2012; Lewellyn et al., 2017). Bergh et al. (2017) measure pathologies using statistical meta-analysis to estimate rates of "unlikely" statistical test results. Park et al. (2017) use meta-analyses of measurements of standard parameters in materials science to track the rates of extreme outlier findings. For the sample of studies with enough replications to estimate reliability of the findings, they estimate about 20% of results in NIST's corpus of findings regarding CO_2 adsorption in metal-organic framework materials are statistically unlikely, although they are careful not to attribute causes or motives to these abberant findings. Salandra (2018) uses expert judgements reported in Cochrane Reviews to estimate clinical trial papers with high risk of bias.

Here, following usage in recent reviews in organization theory, we will use the term pathologies broadly, to include such cases as data fabrication, falsification, plagiarism, significant error, and other

Fig. 1. PubMed retraction trend, 1975–2015.
Note: The dashed line displays the yearly frequency of retraction events in PubMed as a whole, all retraction reasons included (left vertical scale). The solid line displays the yearly percent retracted (right vertical scale), considering two publication types: journal articles and reviews. (Accessed on February 1, 2017).

behaviors that generally fall under the categories of "research misconduct" or "detrimental research practices" (NAS, 2017). For example, Fanelli (2009) conducted a meta-analysis on 18 surveys investigating data fabrication, falsification and modification to improve results ("cooking" the data). His estimate shows that about 2% of respondents admitted to engaging in serious misconduct and about one-third admitted to engaging in lesser forms of deviance. However, as Fanelli admitted, given the sensitive nature of this topic, obtrusive research methods such as surveys are likely to substantially under-report pathologies. Using plagiarism detection software, Honig and Bedi (2012) find 25% of papers accepted at the Academy of Management's International Management division had some evidence of plagiarism and 14% had substantial plagiarism (at least 5% of the text).

As a non-obtrusive measurement, there has been a growing literature on research pathologies measured by retractions in scientific publishing (Azoulay et al., 2015a,b; Furman et al., 2012; Necker, 2014). Much of this literature focuses on the impact of retractions on scientists' careers or on the impact of the retracted papers on other papers by those scientists. For example, Furman et al. (2012) show that citations to retracted papers decline by over 60% after the notice of retraction, compared to a control group. Azoulay et al. (2015a,b) show that even papers that are related to retracted papers (and published before the retraction was announced) suffer a drop in citations, and that the drop is more severe in the case where the retraction was due to "intentional deception", rather than "no sign of intentional deception". This negative effect is greater for more prominent scientists. They also find that there are fewer new articles and fewer new citations in the whole area after a retraction event. Lu et al. (2013) also show that papers related to the retracted paper (by the same author or in the citation network) suffer a citation decline from the retraction. Jin et al. (2013) find the citation decline is larger for less eminent scholars, while eminent scholars face little penalty from the retraction. Mongeon and Larivière (2016) find a significant drop in productivity (often to zero) for those authors who were identified as having committed "fraud". Furthermore, they find that retractions have negative effects on the productivity of "innocent" coauthors on retracted papers, and the effect is stronger if the retraction is due to "fraud" rather than "error". They also find that for fraud retractions, innocent co-authors also suffer a decline in citations to future work (with some evidence of an increase in citations for cases of error retractions). They also find the effect is stronger for first and last authors (likely due to these authors being held more responsible for results in the publication). Thus, while there are some differences (due in part to differences in what is being compared), this work shows that retractions often have significant impacts on the future judgments by scientific peers that sometimes extend beyond the retracted paper.

2.3. Explaining pathologies in science: individual and institutional factors

While there is substantial concern about rising rates of pathologies and the individual-level effects of a pathology event (such as a retraction), there is less work on the causes of such pathologies. Some recent studies have highlighted a variety of individual and cultural level factors that are associated with rates of pathology (Honig and Bedi, 2012; Lewellyn et al., 2017). This work finds that factors such as country of origin, gender and seniority are associated with rates of pathology. Lacetera and Zirulia (2011) use a formal model to describe how the reward and cost structures in publication and peer review facilitate pathologies, showing that the likelihood of pathology is higher for low impact findings and for high status scientists. Steen et al. (2013) find that the time to retraction has decreased (although some of this may be
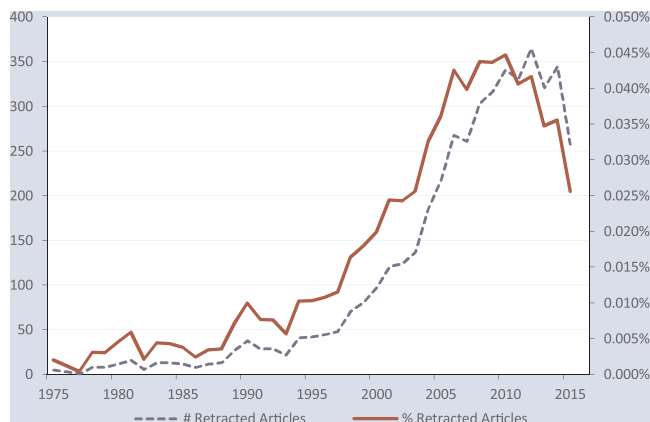
due to truncation bias), and the share of single case retractions has increased. They also find the time to retraction is shorter for high impact journals (suggesting greater scrutiny), consistent with Lacetera and Zirulia's model. Fang et al. (2012) also find that plagiarism and duplicate publication were much more common in low impact journals, while fraud was more common in high impact journals. Salandra (2018) finds that papers with at least one industry co-author have higher risk for selective reporting, as do papers reporting clinical trials for more radical innovations.

Some of this work also points to institutional factors (such as increased competition and demands for increased productivity) that might also produce higher rates of pathologies (Honig et al., 2014). Fang et al. (2012) suggest that the establishment of the Office of Scientific Integrity (now the Office of Research Integrity) and passage of the Whistleblower Protection Act in 1989 may have provided the institutional foundations for an increase in the detection of pathologies in scientific publication. In addition, changes in the incentive structure of science, in particular the diffusion of the New Public Management to the governance of science, with its emphasis on high-stakes evaluation and incentive systems, may have also encouraged scientists to engage in more pathological behaviors (Franzoni et al., 2011; Lewis, 2015; Osterloh and Frey, 2015; Whitley et al., 2007).

There is also evidence for cultural and institutional factors related to cross-national differences in rates of pathologies. Fang et al. (2012) find that the US, Germany, Japan and China together accounted for the bulk of retractions labeled as fraud, while China and India together accounted for more total cases labeled as plagiarism, and duplicate publication, cases than the US. Honig and Bedi (2012) find that rates of plagiarism are higher in non-core countries, arguing that this may be due to the norms about proper research practices being less institutionalized outside the core countries, and that instead there is a loose coupling between publishing as institutional isomorphism and the underlying practices of research (cf. Meyer and Rowan, 1977). Honig et al. (2014) note that as a scientific field becomes more global (using the example of management), the assumptions about a shared set of norms, socialization practices and self-monitoring behaviors become increasingly less tenable, leading to increasing likelihood of pathologies. This would be especially so using a labeling definition of pathologies, as social control agents are likely to be centered in one culture (particularly the United States and some major European countries) while those conducting the research being questioned are increasingly likely to come from another culture (particularly, recently, countries in Asia). Lewellyn et al. (2017) find that researchers who are from countries that score high on corruption scales are more likely to engage in pathologies (conference double dipping). These findings suggest that national differences in tolerance for corruption in business and government may be associated with rates of pathologies in science.

Hypothesis 1. Countries with higher rates of societal (business, government) corruption have higher rates of pathologies in science.

In addition to corruption, incentive systems can also be pathogenic (Larkin and Pierce, 2016). Larkin and Piece (2016) note that the optimal level of pathologies is non-zero, as such pathologies are an outgrowth of effective incentives that motivate increased performance. In the context of science, rewards for excellent research results (for each team member and for the team overall) can produce incentives to fabrication, falsification, or liminal practices that yield desirable findings (such as strategically reporting positive statistical tests and suppressing negative ones). In particular, as the Larkin and Pierce review of the incentives literature points out, the group-based performance metrics of a scientific collaboration might incentivize cheating of various forms by the members, in service of the team's goal of a high-profile result. Larkin and Piece also suggest that high-stakes incentive systems based on performance can lead to excessive risk taking. They give the example of the behaviors of real estate agents, mortgage brokers, and investment bankers during the housing bubble of 2005-2008. In the

context of science, excessive risk taking is likely to produce pathological outcomes due to failing to carefully check or fully report findings or even to fabricating results. Franzoni et al. (2011) discuss the growth of publication-linked incentives in various countries, and show that such publication bonuses are associated with more ritualistic publication practices (increases in submissions without increases in published papers). For example, Quan et al. (2017) report that Chinese universities currently offer significant per paper bonuses, especially for publications in high-impact journals. We conjecture that such practices might also be associated with higher rates of pathologies.

Hypothesis 2. Countries with high-stakes incentive systems have higher rates of pathologies in science.

Thus, this prior work examines individual, cultural and institutional drivers of pathologies in science. Note that most of these prior studies on individual, cultural or institutional level explanations decompose the papers into individual-level measures, for example, assigning a publication or conference submission to each of its authors and then using author characteritstics as the predictors of pathologies. In the next section, we extend this work by taking an organizational perspective and examining the characteristics of the research team.

3. Toward an organizational theory of pathologies in science

In addition to these individual or cultural/institutional level analyses of pathologies in science, we can examine the team structure factors that are likely to affect the levels of pathologies. Merton's strain theory (Merton, 1938) provides the theoretical background for this line of argument. Merton argues that a focus on ends rather than means can produce deviance, especially when there is strong external pressure for achievement and when legitimate means are difficult to pursue.

Merton's discussion of pathologies in science echoes this strain theory:

> "The culture of science is, in this measure, *pathogenic*. …Contentiousness, self-assertive claims, secretiveness lest one be forestalled, *reporting only the data that support an hypothesis*, false charges of plagiarism, even the occasional theft of ideas and *in rare cases, the fabrication of data*,–all these have appeared in the history of science and can be thought of as deviant behavior in response to a *discrepancy between the enormous emphasis in the culture of science upon original discovery and the actual difficulty many scientists experience in making an original discovery*. In this situation of stress, all manner of adaptive behaviors are called into play, some of these being far beyond the mores of science." (p. 323), emphases added (Merton, 1973).

Similarly, in reviewing the sociology of science literature on pathologies in science, Zuckerman (1988) and Hackett (1994) argue that in addition to individual psychology as an explanation, anomie and alienation resulting from the norms and work context of science are additional explanations for pathologies in science. Hackett also includes the problem of excessive span of control and interdisciplinarity making it difficult for supervisors and co-authors to check others' work as additional causes of pathologies. Hackett also highlights the need for a social control perspective for understanding when behaviors are and are not labeled as pathological. Hence, from this perspective, or from a more general sociology of work perspective (Hughes, 1984), it is not surprising that pathologies occur in science (despite a variety of checks in science systems designed to protect against this). The more interesting question then becomes when are such pathologies more or less likely.

Vaughan (1999) and Greve et al. (2010) argue that organizational theory can provide insights for understanding variations in rates of pathologies that move beyond individual or cultural level explanations, and also argue that this area is under-explored in organization theory. Prior work in organization theory argues that complex processes that

are not well understood by other members of the organization can both generate accidents and can provide an environment that can cover active malfeasance, and, furthermore, make it difficult to distinguish which has occurred when an organizational pathology is discovered (Fox and Braxton, 1994; Greve et al., 2010; Paté-Cornell, 1990). Paté-Cornell (1990) expands engineering models of system failure that focus on external events and system components to incorporate organizational and managerial causes of failures that lead to disasters. Goodman et al. (2011) make a similar argument from an organizational behavior perspective, arguing that organizational pathologies should be understood as organization-level phenomena, which are not reducible to individual motives or intents. In their review of organization theory and pathologies in science, Baron et al. (2016) suggest key organizational factors include: organizational design, demography, and the division of labor; network embeddedness; institutional fields; and status processes. Vaughan (1999) argues as well that complexity and the division of labor are key sources of pathologies in organizations. She further argues that the nature of the task, in particular, high uncertainty, can compound the rates of pathologies. Division of labor can create problems of coordination and oversight (Blau, 1970). Paté-Cornell (1990) argues that strict compartmentalization of tasks that lacks sufficient feedback or cross-cutting horizontal communication channels is likely to result in information gaps and weak risk management, allowing both errors and malfeasance to escape notice, leading to system failure. In their discussion of the social structure of a conspiracy, Baker and Faulkner (1993) show that illicit activity is easier to maintain in the face of a sparse network structure. In particular, if each person works in relative isolation, only connected to people who are unconnected to each other, then it is difficult to observe and verify that illicit activity has happened. Goodman et al. (2011) argue that division of labor may also make it more difficult to spot the problems or to respond appropriately to ameliorate them before they have adverse consequences for the organization.

Greve et al. (2010) note that organizational structures are designed to economize on decisionmaking, with division of labor and standardized routines implemented to increase efficiency in the face of stable task expectations. However, in the face of exceptions, such specialization and routinization can produce suboptimal outcomes. Similarly, Riemer (1976) shows that the high degree of division of labor and specialized skills is one major source of "mistakes" (deviations from the plans) in building construction. Furthermore, these mistakes are cumulative, such that the mistakes by one skilled tradesman become the foundation conditions on which the next tradesman builds, in turn adding his own mistakes. Hence, a completed building is often permeated with imperfections. Jones (2009) argues that as the burden of knowledge in technical fields increases, there is greater need for specialization and division of labor in a collaboration. However, here we argue that one outcome of that burden of knowledge is an increase in pathologies in science.

Latour and Woolgar describe the division of labor in the biology lab they study: "Individuals referred to as doctors read and write in offices in section A while other staff, known as technicians, spend most their time handling equipment in section B. … Section B appears to comprise two quite separate wings: in the wing referred to by participants as the "physiology side" there are both animals and apparatus: in the "chemistry" side there are no animals. *The people from one wing rarely go into the other*." (Latour and Woolgar, 1979: 45, emphasis added). As Latour and Woolgar (1979) show, information flows within a scientific collaboration involve a series of transformations and simplifications as they move from raw observations to scientific findings. This division of labor may be a key source of either unwitting errors or deliberate malfeasance that can lead to retractions. Paté-Cornell (1990) emphasizes that such systems may fail to communicate the uncertainties and reservations that accompany the reported information. This can lead to over-reliance on the information as it is processed and reprocessed, often in increasingly abstract forms (Latour and Woolgar, 1979), which

may be a key source of error, and what might even get constructed as malfeasance, in scientific teams with extensive division of labor. As the quote highlights, this condition can hold even in local collaborations, i.e., it is not synonymous with remote collaboration. The practice of outsourcing certain tasks, for example: testing samples, statistical analyses, or collecting clinical data, may compound the division of labor problems by further removing the oversight of, capacity to do, understand and evaluate, and the sense of responsibility for, the particular piece of the research coming from another organization (Baron et al., 2016). In clinical trials, for example, the contract may call for generating a quota of patients, which can lead to pressure on the clinic to produce the data, by any means necessary, and can lead to weak oversight by the lead investigators, who may not have any ready way to check the veracity of the submitted data (Fisher, 2009). The existence of the specialty of clinical trial managers, who coordinate the various components of a disparate team (study designers, clinics collecting data, biostatisticians, research faculty, technicians, etc.), may create a sparse network structure that may further exacerbate these problems (Fisher, 2009; Sismondo, 2009).

Furthermore, organizations, in practice, tend to develop over time, innovating to improve performance and to address localized problems, and the accretion of such deviations from established procedures can, over time, produce significant drift from formalized practices (Halle, 1984; Zuboff, 1988). These modifications learned over time may work very well as long as the variations in the task environment fall within normal bounds, but may be vulnerable to breakdowns or mishandling of exceptions in cases where inputs are not as expected (Greve et al., 2010). Riemer (1976) argues that the interdependent yet specialized and distinct tasks of the tradesmen on a building site along with the presence of a variety of exceptions not anticipated in the plans require a process of negotiation to jointly accomplish their individual tasks in this shared space, and that these negotiations often require adaptations, modifications, and changing of details that lead to pathologies. Similarly, in the case of scientific collaborations, the different task specialties in the project (for example, those collecting and preparing samples, those running assays on the samples and those doing the statistical analyses of the resulting data) may each view the requirements of the inputs and the outputs of these processes slightly differently, and the result may be modifications, errors and even falsifications in order to satisfy the jointly negotiated requirements imposed on each project member (Roth, 1966). Note that this problem is inherent even if all those performing the distinct task are from the same discipline, i.e., this is not synonymous with interdisciplinarity. Paté-Cornell (1990) refers to these various deviations introduced through division of labor as "resident pathogens", waiting to express themselves as organizational pathologies. In addition, Riemer (1976) notes a relation between productivity pressures and pathologies at work, often leading to wasted materials or the need to redo parts of the work. Riemer finds that such "circumstantial errors" are especially likely to occur near the end of a project, as deadline pressures push workers to trade craftsmanship for speed. Honig et al., (2014) note that journals imposing short turn-around times for revisions can exacerbate the time crunch pressures that lead to shortcuts (or mistakes) in the science. Vaughan (1999) argues that this division of labor and specialization segregates knowledge about each participant's tasks and goals, creates "structural secrecy" and increases the likelihood of errors as information passes across boundaries from one segment to another.

In addition, Merton (1973: 332) and Hackett (1994) note that the growth of collaboration and division of labor in science may lead to an alienation of the scientist from the scientific research project in which she is but a part. Furthermore, Hughes (1984) argues that in dependent occupations (such as the nurse and the pharmacist with reference to the physician, or the technician or graduate student with reference to the senior faculty) there may be especially strong adherence to the rituals of procedure, perhaps to the detriment of substantive performance, as a protection against the possible mistakes of the physician or the senior

scientist. The bureaucratization of scientific work, including the extensive division of labor between principle scientists and data collectors and the routinization of data collection, can generate a "hired-hand" attitude toward the work and produce regularized pathologies (Roth, 1966). As was the case as manufacturing work transitioned from craft to industrial production, the separation of conception from execution can reduce commitment to the outcomes of the work (Braverman, 1974; Chinoy, 1955). This alienation from the goals of the project can further increase the risk of pathologies as the immediate objective becomes to complete assigned tasks in order to fulfill what may come to be viewed as the external demands of a role (rather than the joint production of common output).

Furthermore, Hughes (1984) argues that occupational groups will jealously guard the right to determine correct procedures and standards of output for their occupation, especially in cases where it may be difficult to objectively determine performance. Chubin and Hackett (1990) make similar arguments about scientists. In addition, in a complex scientific project, responsibilities may diffuse and there may be an assumption that the other members will have checked and rechecked the equipment, samples, data and analyses under their perview, so that problems can "fall through the cracks" (Baron et al., 2016). At the same time, a logic of good faith is likely to prevail among professionals, which increases the likelihood that interim results from colleagues will be taken as correct, making any pathologies more likely to get incorporated into the final product (Chubin, 1985; Honig and Bedi, 2012). The Bell Labs report on the Hendrik Schön case notes the difficulties with confronting colleagues and violating this norm of trust, while at the same time suggesting that at least the senior scientist on the team (Batlogg) should have crossed this line and questioned Schön about some of the results (Beasley et al., 2002).

In addition, Hughes (1984) argues that one of the functions of systems of roles in a division of labor is the delegation of the risk of and guilt for mistakes and the associated losses due to them, as the examples of David Baltimore or Donald Green illustrate. Hence, one argument for the likely relation between division of labor and retractions is that the ability to focus the blame on one or another subcomponent of the team if such pathologies come to light may reduce the insistence on the control, supervision or redundancy that would be needed to prevent such pathologies, especially if such control comes at the perceived cost of reduced productivity. Larivière et al. (2016) argue that the rise of contribution statements that specify which authors are responsible for which tasks might reduce rates of pathologies by fixing responsibilities on particular named individuals. In contrast, we are arguing that such statements would increase pathologies by allowing authors to deny responsibility for some of the tasks in the project. In fact, one interpretation of Mongeon and Larivière's (2016) or Jin et al.'s (Jin et al., 2013) data on differential impacts from retraction is that the scientific community is implicitly assuming a division of labor, and assigning differential responsibility for the pathological outcome. Division of labor may reduce accountability, especially to the extent that senior people are more immune from adverse effects. Hence, division of labor in science can create conditions for rationalization or disengagement leading to pathologies. Larivière et al. (2016) note that the increasingly common contributorship statements can be used as a window into the division of labor in a project. We will take advantage of such contributorship statements to examine the relations between division of labor and pathologies. Larivière et al. (2016) also conjecture that specifying the contributions of each author in the presence of division of labor might reduce rates of pathologies, although they note that it is an open question. We will test this, arguing that literature from organization theory suggests pathologies should increase in the face of division of labor.

These arguments suggest that these pathologies are: more likely to occur, less likely to be found out and corrected, and more likely to cumulate in work organizations characterized by a strict division of labor among highly skilled specialists working in a high uncertainty environment. Hence, there are a variety of mechanisms by which division of labor can increase the likelihood of pathologies, including structural secrecy and miscommunication, alienation/goal displacement and delegation of responsibility. Organizational-level pathologies are social facts, *sui generis* (Durkheim, 1951; Goodman et al., 2011), and cannot be explained by recourse to individual-level explanations. Hence, we will move beyond the individual level and examine the extent to which organizational structures themselves are pathogenic (Merton, 1973; Paté-Cornell, 1990; Goodman et al., 2011; Baron et al., 2016). In particular, we will focus on the relation between the division of labor and the rates of pathologies in science.[1]

Hypothesis 3. Greater division of labor is associated with higher rates of pathologies in science.

In addition to these structural and contextual factors, there may be aspects of the research topic itself that increase the rates of pathologies, such as novelty or complexity. We will use a variety of methods to control for these alternative potential causes, such as matching (see below), as well as controlling for interdisciplinarity, field, journal impact factor, university rank, and publication year. However, given the observational nature of our data, we cannot rule out all potential confounding factors related to unmeasured heterogeneity across projects. Hence, we will report empirical findings that are consistent with our theoretical predictions, but that cannot definitively establish causality.

4. Data and measures

To test these hypotheses, we need data on a large number of papers, some of which were retracted, as well as data on the structure of the collaboration. We examine division of labor and retractions in biomedical and life sciences, where retractions tend to be sufficiently common to allow us to analyze these relationships (Lu et al., 2013; Tang et al., 2016). In addition, in these fields it is becoming increasingly common for the published paper to include contributorship statements, which can be used to estimate the division of labor (Larivière et al., 2016).

Our data were constructed in the following sequential steps. We first retrieved retracted articles indexed in PubMed Central® (PMC), published since January 1, 1975. We used PubMed Central because this database allows free access to the full texts of articles, which gives us access to contribution statements. We limited the retracted document type to reviews or articles. The search was completed on March 20, 2016 and returned 1081 hits. As the information on authors' contributions for the paper is disclosed in different sections of PMC papers, we had two researchers read the full texts of the 1081 papers and cross check for consistency in identifying the contribution information. About 29% of the papers have the relevant information.[2] However, further checks showed that some of the contribution statements were primarily about authorship order, rather than task division, and hence were eliminated (since they give no information on the division of labor). Furthermore, we limited papers to those that were also listed in Web of Science, in order to collect supplemental data on control variables.

When we compared the retracted papers from PMC to the retracted papers from the entire PubMed database (excluding those from PMC), limiting to cases with at least 2 authors, the retracted papers from PMC are not significantly different on number of authors, but are, on

---

[1] Some may argue that scientific collaboration is a self-governing or self-organizing network, not formed by structural or institutional factors (Wagner and Leydesdorff, 2005). However, we argue that even in a self-organizing collaboration, if the collaboration produces division of labor, the risk of pathologies gets higher.

[2] The section headings in which contribution information appears include: Footnotes, Authorship, Author contributions, Author information, Authors' contributions, Authorship Contributions, as well as Acknowledgement/ Acknowledgments/Acknowledgements.

average, more recent (the average of publication year is 2008 v. 2007 p < .001), more multi-organizational (1.2 versus 1.1 organizations, p < .001) and more likely to report funding (31% v. 15%, p < .001). When we control for publication year, only the difference in reporting funding is significant. Hence, aside from differentially reporting information on funding sources, the retracted papers in the PMC database (controlling for publication year) are representative of the overall PubMed database on, for example, number of authors or number of organizations per paper.

Within the set of retracted papers in PMC with 2 or more authors and also listed in Web of Science, when we compared the retracted papers with contribution statements to those without contribution statements, we see those with contributions statements are generated by larger (6.6 versus 5.6 authors, on average, p < .001), and more multi-organizational (2.8 versus 2.0 organizations, p < .001), teams, although, again, the differences are modest, though statistically significant. For example, the difference in team sizes is about 1/3 of a standard deviation. We also see that those with division of labor statements are published more recently (mean publication year, 2011 v. 2005, p < .001), as well as having fewer citations (24 v. 35, p < .001) and being more likely to report funding (58% v. 41%, p < .001). Since some of these differences may be due to differences in publication year (as work practices and reporting practices have changed over time), we re-ran our inter-group comparisons controlling for year. After controlling for year, we find that those reporting contributorship information have more authors on average, and more affiliations, but other measures (funding reported, citations and references) are not significantly different between the two groups (PMC retracted papers with or without contributorship statements). Later, in our results section, we will control for publication year and number of authors, and check for the robustness of our findings using number of affiliations. Still, it should be kept in mind that, on average, our data (PMC retracted papers with contributorship information) represent somewhat larger than average teams and somewhat more recent publications compared to the whole population of PMC retracted papers.

Next, to develop a comparison set of papers, we searched adjacent papers for these PMC articles with contributorship statements and WoS information, for up to two non-retracted articles, one immediately before and one immediately after the retracted paper, in the same issue of the same journal and including contributorhip information and a listing in the Web of Science, similar to Azoulay et al. (2015b) and Furman et al. (2012). If neither of the adjacent papers is qualified (i.e., if the document type is correction, editorial, letter, and so on), we extended the selection further with a maximum searching up to three papers ahead of or behind the targeted retracted one. If the retracted article was the lead paper or the last one in that issue, only one comparison article was included.

Note that unlike prior studies that use retraction or not as being "treated" or not (e.g., Azoulay et al., 2015a,b; Furman et al., 2012), we use retraction or not as a dependent variable. Using the comparable sets of retracted and non-retracted papers helps develop a sample of cases that are similar on research domain, publication year and journal ranking, thereby reducing unmeasured heterogeneity of the comparison sets. Using this sample of papers, we would like to see how structural characteristics such as division of labor are associated with retraction. In such observational data, a concern affecting inference about effects of the treatment may be that selection (by agencies or participants) into the "treatment" (in our case, division or labor or not) may be biased in two ways, either one group is just better overall on the dependent variable or they benefit more from the treatment (Morgan and Winship, 2012). In our case, the "treatment" is division of labor (not retraction as is the more common case in studies of the effects of retraction). We have no reason to believe that teams choose division of labor to increase retraction (selecting on the dependent variable) nor that the teams that adopt division of labor are especially sensitive to producing retracted papers due to division of labor. In fact, we would expect the exact

opposite (i.e., any selection would be motivated by a desire to reduce retractions).

After validation and cleaning, and dropping duplicates and solo authored papers, as well as dropping a small number of papers that were retracted due to publication errors, and limiting to cases that have complete information on our key predictor variables (division of labor, corruption and incentives), our dataset contains information on 544 papers (i.e., 195 retracted and 349 comparison) for analysis, covering publication years from 2004 to 2016.

4.1. Dependent variable: retraction

The dependent variable is a dummy variable *retraction*, which is coded 1 if in the retracted group and 0 if in the comparison group. While this widely used measure likely under-estimates the total rate of pathologies (broadly defined), it is an important measure in its own right, as it is a marker of cases where the pathology was eventually discovered and where the problem was constructed to have reached a level of seriousness sufficient enough to require a public statement of the pathology.[3] Furthermore, it means that the result has been nullified, removing it from the body of scientific knowledge. Hence, this operational definition of pathology reflects Greve et al.'s definition of pathology based on Becker's labeling theory (Becker, 1963; Greve et al., 2010).

4.2. Division of labor

We develop a division of labor measure that represents the structure of the research team. Division of labor (DoL) can be conceptualized, and operationalized, in a variety of ways. For example, using PLOS publications, Larivière et al. (2016) and Macaluso et al. (2016) study the relations between tasks and authorship, and how these vary by author characteristics (seniority, gender) and field. Larivière, et al. find, for example, that senior authors are more concentrated in cognitive tasks (such as writing or conceptualizing) while junior people are more concentrated in technical tasks (such as data collection and analysis). They also find that there is more division of labor in medicine than in mathematics, physics and social sciences. In their case, since their focus is on the span of tasks of each individual in the team, they operationalize division of labor as number of tasks a given person does.

Our division of labor measure focuses on *task* partitioning and redundancy to represent how likely the *task* is to be reviewed by more than one member. It represents the levels of shared understanding of the task and quality checks in the team. Prior work in organization theory suggests that such redundancy may be important for limiting mistakes (Roberts and Libuser, 1993). Similarly, Baker and Faulkner's discussion of the social structure of a conspiracy suggests that a sparse network (little redundancy) makes it more likely to engage in

---

[3] In this study, distinctions between what might be called "intentional malfeasance" versus "unintentional errors" versus "plagiarism" is not the main focus, as we are looking at when we are likely to see more or fewer pathologies of any kind. Based on our theoretical arguments, the greater the division of labor, the more likely the project is to generate intentional malfeasance, plagiarism or unintentional errors. Which kind of pathology is more or less likely to happen as the division of labor increases is not predicted by the theory we are using. Moreover, when we empirically tested our model distinguishing intentional malfeasance from unintentional errors from plagiarism using data on reasons for retraction, we did not see any significant deviations from our main findings. When we compare the relation between division of labor and retraction separately for intentional malfeasance, unintentional error and plagiarism (using the codes from RetractionWatch), the relation is positive in all three cases, and similar to the original estimate for all retractions, although significance levels change, likely due to modest sample sizes in these subsample regressions. We do not find a statistically significant difference in the division of labor coefficients across the three types of pathologies.
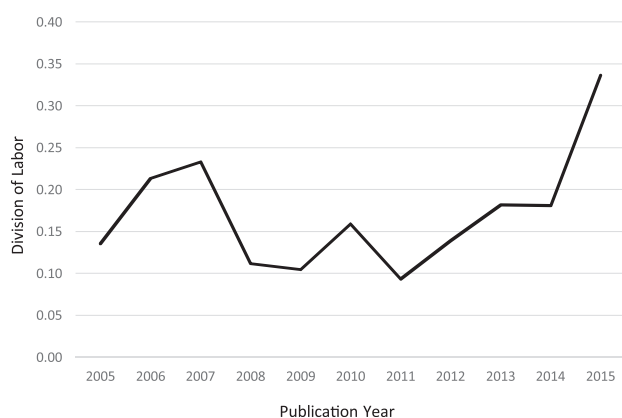
Fig. 2. Mean division of labor, by year, 2005–2015.

questionable behaviors without being caught. Hence, low redundancy division of labor should be more likely to produce pathologies (Baker and Faulkner, 1993). Furthermore, building on the recommendations from NIH's investigations of a high-profile case of pathologies, Chubin (1985) recommends task redundancy to reduce pathologies in science, suggesting that division of labor (in the form of lack of task redundancy) is a likely cause of pathologies. Therefore, our measure of division of labor reflects the theoretical arguments about the links between division of labor and pathologies, at the research team (rather than individual) level.[4]

To create this measure, we use contribution statements for each paper in our sample (Corrêa et al., 2017; Haeussler and Sauermann, 2016; Larivière et al., 2016). Prior studies that use contribution statements in journals assigned authors into one or more of these six tasks: 1) Conceived and designed the study, 2) Performed the experiments, 3) Contributed materials/research tools, 4) Analyzed the data, 5) Wrote the paper, and 6) Other. These prior studies limit data to PLOS journals, which provide more standardized descriptions of authors' contributions, making it relatively straightforward to assign authors into one of the six tasks (Haeussler and Sauermann, 2016; Larivière et al., 2016). However, PubMed Central includes papers from various journals in biomedical and life sciences (as well as PLOS journals) where some show more standardized descriptions of authors' contributions while others have various contribution statements. Therefore, still adopting those task categories, we first go through each contribution statement to assign authors to one or more of the six tasks (See Tables A1 and B1 in Appendices A and B). Some papers have authors spread over all six categories whereas others cover only some of these six categories. Therefore, the number of divided tasks is not the same for all papers in our sample.[5]

Next, for each paper, we calculated the share of tasks that were performed by at least two members. Then, we computed 1 minus the share of tasks that were performed by at least two members to make a higher value represent more division of labor (task isolation). Therefore, our measure of division of labor shows the share of tasks that were performed by only one member (see Eq. (1) and for a detailed description of this process, with some examples, see Table B1 in Appendix B). One advantage of this single person per task measure is that

it is measuring the extent to which there is a clear assignment of responsibility (cf. Biagioli, 2003; Larivière et al., 2016). Hence this measure may also help distinguish between "responsibility/visibility" of authors as a driver, which would suggest a negative relationship with retractions (Larivière et al., 2016), versus a model based on organizational theories related to structural secrecy, goal displacement and alientation, and work-arounds combined with a norm of good faith, which would suggest a positive relationship between division of labor and retractions (Riemer, 1976; Vaughan, 1999). Hence, this measure, if it is in the hypothesized positive direction, more clearly rules out some possible rival explanations for the relationships between division of labor and pathologies. On average 16% of tasks in a team are performed by only one member (See Table 1). If we examine each task separately, we find that the task (excluding "other") that is mostly likely to be solo is providing materials/data, with about 26% of the cases that report that task reporting that it is solo. The rest of the tasks are all 15% ± 1% solo (among those who report that task). We also examined the trends in the division of labor over time, limiting to publication years having at least 10 observations. Fig. 2 shows, from 2005 to 2015, overall, a fairly flat trend in the division of labor.[6]

$$DoL = 1 - \frac{\text{\# of tasks perfomed by at least two members}}{\text{Total \# of tasks presented in contribution}} \quad (1)$$

This measure may suffer from certain measurement errors. The first is that, even within categories, there may be specialization, such that, for example, one person does one part of an experiment and another does a different part. In this case, our measure will underestimate the degree of division of labor (downward bias). Such a bias would likely bias our regression coefficients toward zero, giving us a conservative test of our hypothesis 3. A second bias is there may be a social norm towards reporting that all team members participated in all of the tasks. Or, similarly, there may be a norm to overly share credit for specific tasks (for example, writing, as shown in Larivière et al., 2016). In our data, 55% of teams have no division of labor (i.e., all tasks include at least two members). Third, it is possible that part of the work was outsourced to external technicians but those technicians were not included in the author list, which contributes to underestimating division of labor measured by contribution statements (Jabbehdari and Walsh, 2017; Walsh and Lee, 2015). Again, these sources of error would bias our measure toward zero (the measured division of labor would be lower than the actual level of specialization). Accordingly, this would dampen any observed effects, giving a conservative test of the hypothesis.[7]

4.3. Institutional context

In addition to our measure of division of labor, based on the arguments above, we also test the effects of institutional context on the rates of retraction. We have two country-level variables for institutional predictors: 1) the level of corruption, and 2) research incentive systems.

---

[4] Bunderson and Sutcliffe (2002) show that different conceptualizations of functional diversity can generate very different results, often not being comparable, and hence suggest that researchers should explicitly define measures that match their theories.

[5] Starting from our sample of retracted and matched papers, for the division of labor measure, we had to do extra checks on each contribution statement because some contribution statements say, for example, "These authors contributed equally" and this statement is related to authorship, but does not provide any information on task distribution among authors. We treat these cases as missing in the division of labor measure.

[6] A linear regression line for these data show a slope of .01, t = 1.03, p = .33.

[7] One additional source of error is coding error. There is some amount of judgment involved in assigning non-standardized text to the particular tasks in our coding scheme. The coding rubrics were developed collectively between two of the authors. Once the coding scheme had been established, the division of labor data were coded by one author for all cases. Then, to check the reliability of this coding, a random sample of 10% of the cases was independently coded by another author (cf. Honig and Bedi, 2012). The result was 78% overlap in the division of labor score for the two codings, with a correlation of .83 across the two sets of scores. Hence, we have confidence that the coding is within acceptable levels of reliability. However, to the extent that there is error in the coding, this is likely to dampen any effects of division of labor on retraction, again giving us a conservative test of the hypothesized relation.

4.3.1. Corruption

For the different levels of corruption among countries, following Lewellyn et al. (2017), we used the Control of Corruption indicator by the World Bank, part of its Worldwide Governance Indicators (World Bank, 2016). The corruption indicator measures "perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests" (World Bank, 2016). The indicator ranges from -2.5 (weak) to 2.5 (strong) governance performance, where stronger governance means lower control by private interests, showing a lower level of corruption. Therefore, we reverse-coded the values of the indicator. Then, for each team, we checked the corruption scores associated with each different country represented in the author list and used the maximum corruption score. We use the 2000 survey for our measure. When we compare between the 2000 and 2015 Control of Corruption indicators among countries, we find that the correlation in country scores across the two surveys is .92. Similarly, we calculated the maximum corruption score for each paper using the 2000 and 2015 indicators and find these two measures are correlated .96. These findings suggest that this corruption indicator is fairly stable, and not sensitive to the year chosen. We also used a mean corruption score for robustness checks.

4.3.2. Incentives

Following on Franzoni et al. (2011), we measure incentive systems by classifying countries into the ones with any incentive for international publication, including institutional incentives, individual career incentives or individual cash bonuses (11 countires) contrasting with countries without such incentives (19 countries).[8] For each team, we checked each member's country and assigned 1 for the team where at least one member is from a country with an incentive system, thereby generating a binary variable for any incentives. We also used whether at least one member's country has cash bonuses or not (i.e., China, Korea and Turkey vs. others) for robustness checks.

4.4. Controls

4.4.1. Number of authors

Brass et al. (1998) argue that as the size of the group increases, the increase in surveillance, and the greater risks to reputation, should reduce pathologies, although these effects are likely conditioned on the structure of relationships among team members. At the same time, larger team size may make monitoring more difficult, increasing the likelihood of pathologies (Nagaoka and Owan, 2014). Moreover, our division of labor measure may be affected by team size, because the chance to share a task may go up with larger teams. Therefore, to see the net effect of division of labor, we control for number of authors. We count all authors, collected from WoS, where the minimum is 2 and the maximum is 56. We also used log number of authors for a robustness test.

4.4.2. Interdisciplinarity

One additional possible confounding factor in the division of labor to retraction link is interdisciplinarity. It may be the case that interdisciplinarity may generate greater scientific difficulty, thereby increasing the chance of making mistakes, either simple errors or differing perspectives on judgment calls, that might lead to retractions. It may also be the case that interdisciplinary projects receive greater scrutiny from the scientific community and hence are more likely to be retracted. Furthermore, division of labor may be associated with greater interdisciplinarity (as one reason for division of labor might be bringing together specialists from different disciplines). However, our meaning of division of labor and interdisciplinarity do not have to be related, as, for example, a team of cell biologists who divide up the work between designing an experiment, extracting materials from the cells, doing assays on the materials, analyzing the results and writing the paper might have a very high division of labor, but all drawing from the literature in cell biology, and hence not being interdisciplinary. Therefore, we control for interdisciplinarity of research projects, using the count of different mega disciplines cited in the paper (Porter and Rafols, 2009).

To create this measure, we first retrieved all references of each paper and generated Cited Subject Categories (SCs) based on the journals cited in the focal paper. The Web of Science (WoS) has over 200 SCs. We next aggregated the SCs into 21 mega research fields using the approach developed by Porter and Rafols (2009) and Tang et al. (2015). The count of different cited mega research fields is then computed as a proxy measure for interdisciplinarity. In our data, in fact, the interdisciplinarity variable is not correlated with division of labor ($r = -.08$), suggesting that the relation between division of labor and retraction is not likely due to confounding effects from interdisciplinarity. However, we still include interdiscplinarity as a control, to correct for the potential omitted variable problem, and also to test the direct effect of interdisciplinarity on retractions, which may be of interest in its own right.

4.4.3. Competition (high-stakes paper)

We also control for the impact factor of the journal in which the paper is published, as a control for competitive pressures. First, papers published in prestigious journals on average are exposed to scrutiny by a wider scientific community (Lacetera and Zirulia, 2011). Furthermore, such papers may proxy for the level of competition faced by the research team (as the pressure to publish quickly is likely highest in premier journals). For example, Vaughan (1999) highlights the role that competitive pressures from the environment have on pathologies at work. In an experimental study Belot and Schröder (2013) find that both productive and counterproductive (pathological) behaviors increase under conditions of competitive reward structures (see also Larkin and Pierce, 2016). Similarly, Paté-Cornell (1990) argues that time pressure can lead to accepting minor flaws if they do not cause any immediate problems, although such flaws can later produce (sometimes catastrophic) system failures. Hence, we may need to control for levels of competition faced by the research teams, with journal impact factor as a proxy. Previous studies have shown that cases of retractions are positively related to journal impact (Noyori and Richmond, 2013; Steen, 2010). To distinguish high-stakes paper, we use the indicator of field-normalized Journal Impact Factor (JIF) quartile (Liu et al., 2016). For this variable, we retrieved journal impact factors from the 2016 Journal Citation Report. If the paper was published in a journal with a JIF in the top quartile of a specific research domain, it is designated as 1, otherwise 0. If the paper belongs to more than one WoS category, we use the optimistic approach, i.e., its highest ranking.

4.4.4. University ranking

We also control for institution status. Previous studies have shown that organizational norms and characteristics also impact individual's attitude, belief and behavior. Papers from high status institutions may have a higher probability of being retracted (Furman et al., 2012), perhaps because they are more heavily scrutinized, or because these scientists are more socialized into and committed to scientific norms. It is also possible that scholars in top universities care more about their reputation and are less likely to endanger it by submitting a

---

[8] According to Franzoni et al. (2011), the 11 countries where reward policies were introduced are: New Zealand, Australia, Belgium, Norway, Denmark, Italy (for institutional incentives), Spain, Germany (for individual career incentives), China, Korea, Turkey (for individual cash bonuses). The other 19 countries are Austria, Canada, Finland, France, Greece, Hungary, Iceland, Ireland, Israel, Japan, Netherlands, Poland, Portugal, Russia, Singaore, Sweden, Switzerland, UK and USA. If no author was on the combined list of 30 countries, the measure was set to missing.

questionable publication (and hence would have lower rates of pathologies). To control for such effects, we use the 2016 Academic Rankings of World Universities released by Shanghai Jiao Tong University to control for the impact of organization rank.[9] For this variable, we measure whether the corresponding author or reprint author is affiliated with an elite university or not as a proxy indicator of organization culture. We create an ordinal variable, with a value of 5 for ranking 1–25, 4 for 26–50, 3 for 51–75, 2 for 76–100 and 1 for the rest (including those not ranked). As a robustness check, we tested a model using the CWTS Leiden ranking as an alternative measure and find similar results for our hypothesized relations.[10]

4.4.5. Industry affiliation

There are also concerns that pathologies may be higher in the case of industry-affiliated scientists (Salandra, 2018). In addition, it is possible that industry affiliation may be associated with greater division of labor (either within the firm or between the university and industry affliated authors). To control for this, we include a dummy variable if any of the authors has an industry affliation. As a robustness check, we also test if the paper reports any industry funding.

4.4.6. Others

We control for publication year. And we control for field. We start with WoS categories of fields (more than 70 detailed categories in biomedical and life-sciences) and collapse them into 10 categories (i.e., Biochemistry, Biotechnology, Cell biology, Biology nec., Medicine, Microbiology, Neurosciences, Pharmacology, Virology, and Other). We use dummy variables to control for these 10 fields, with "Other" as the reference field.

5. Results

5.1. Descriptive statistics

Table 1 shows the descriptive statistics and correlations. We find that there is a negative correlation between team size and our measure of division of labor. This may be a result of the limitations of our measure of division of labor noted above. In particular, since we have only six categories, we may not capture the detailed divisions within, for example, doing the experiment. Moreover, since our division of labor measure captures the share of tasks that were performed by only one member (to focus on the level of monitoring, coordination and redundancy), this may be less likely to be observed in larger teams, especially with a fixed set of six tasks categories. Therefore, we test the effect of division of labor controlling for size, which helps address this, showing the effect of division of labor independent of size. Table 2 compares retracted papers to non-retracted comparison papers for our variables, and shows that the two groups are equivalent for team size, competition (measured by journal ranking), university ranking, industry affiliation, research domain and publication year, suggesting that the comparison sets are very similar (reducing concerns about unmeasured heterogeneity explaining our results). However, we see differences in division of labor and institution variables (corruption and incentive) between retracted and non-retracted papers (which suggests that these variables may be related to pathologies). For example, the mean division of labor (percent of tasks that are solo) is 19% for the retracted paper sample and 14% for the non-retracted papers (p = .01). Hence, among retracted papers, there is a 36% higher chance of a given task being performed solo.

Table 1
Descriptive statistics.

	Variable	N	Mean	SD	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Retraction	544	0.36	0.48	0	1	1.00															
2	Division of labor	544	0.16	0.21	0	1	**0.11**	1.00														
3	Yes/No	544	0.45	0.50	0	1	0.07	**0.83**	1.00													
4	75th pctl cut	544	0.21	0.40	0	1	0.06	**0.83**	**0.56**	1.00												
5	95th pctl cut	544	0.03	0.18	0	1	**0.15**	**0.56**	**0.21**	**0.37**	1.00											
6	-Avg. members per task	544	-3.44	1.79	-15	-1	0.08	**0.48**	**0.45**	**0.38**	**0.22**	1.00										
7	# of tasks	544	4.10	0.87	1	6	0.02	-0.01	0.08	**-0.14**	-0.05	-0.03	1.00									
8	Conception distinct	499	0.09	0.29	0	1	-0.03	**0.19**	**0.12**	**0.16**	**0.17**	**0.12**	**-0.17**	1.00								
9	# diff. institutions	544	2.45	1.66	1	20	-0.08	**-0.19**	**-0.18**	**-0.13**	**-0.11**	**-0.42**	**0.20**	**-0.10**	1.00							
10	World Bank Corruption Index	544	-1.03	0.92	-2.59	1.01	**0.18**	**0.12**	0.08	**0.09**	**0.15**	0.07	-0.01	**0.15**	-0.08	1.00						
11	Country with pub. incentives	544	0.51	0.50	0	1	**0.13**	0.06	0.01	0.08	**0.11**	0.00	-0.06	**0.10**	0.03	**0.69**	1.00					
12	No. of authors	544	6.84	4.37	2	56	-0.05	**-0.31**	**-0.27**	**-0.27**	**-0.16**	**-0.72**	**0.23**	0.00	**0.10**	0.04	**0.09**	1.00				
13	Interdisciplinary	544	3.53	1.41	1	9	0.08	-0.08	-0.05	**-0.09**	-0.04	-0.02	-0.01	-0.04	-0.08	0.00	0.05	-0.02	1.00			
14	Top quartile JIF	544	0.66	0.47	0	1	-0.03	**-0.16**	-0.08	**-0.18**	**-0.14**	**-0.10**	**0.38**	**-0.11**	**0.15**	**-0.34**	**-0.30**	**0.11**	0.08	1.00		
15	University ranking	544	1.67	1.36	1	5	-0.02	-0.05	-0.04	-0.05	-0.01	-0.03	**0.10**	**-0.10**	0.01	**-0.30**	**-0.30**	0.05	-0.05	**0.28**	1.00	
16	Industry affiliation	544	0.06	0.24	0	1	0.06	0.06	0.03	0.05	0.05	0.01	0.08	0.00	**0.22**	-0.06	0.04	**0.15**	0.07	**0.10**	-0.04	1.00
17	Publication year	544	2011	2.58	2004	2016	0.02	0.06	0.03	0.05	0.05	0.01	-0.08	**0.14**	-0.04	**0.30**	**0.23**	0.04	0.06	**-0.32**	**-0.19**	-0.08

Note: Correlations with bold numbers statistically significant at p < .05.

Table 2
Comparison between retracted papers and non-retracted comparison papers.

Variable	Retracted (N = 195)	Non-retracted (N = 349)	t	p
Division of labor	0.19	0.14	2.46	0.01
Yes/No	0.50	0.43	1.70	0.09
75th cut	0.24	0.19	1.52	0.13
95th cut	0.07	0.01	2.93	0.00
-Avg. members per task	−3.24	−3.55	1.99	0.05
# of tasks	4.12	4.09	0.40	0.69
Conception distinct	0.08	0.10	−0.62	0.53
# diff. institutions	2.27	2.56	−2.13	0.03
World Bank Corruption Index	−0.81	−1.15	4.15	0.00
Country with pub. incentives	0.59	0.46	2.96	0.00
No. of authors	6.57	6.99	−1.09	0.28
Interdisciplinarity	3.67	3.44	1.81	0.07
Top quartile JIF	0.65	0.67	−0.64	0.52
University ranking	1.62	1.69	−0.58	0.56
Industry affiliation	0.05	0.06	−0.56	0.58
Publication year	2011	2011	0.56	0.58
Field				
Biochemistry	0.05	0.04	0.75	0.46
Biotechnology	0.02	0.02	−0.18	0.85
Cell biology	0.01	0.01	−0.13	0.90
Biology, nec	0.04	0.03	0.46	0.64
Medicine	0.37	0.35	0.53	0.60
Microbiology	0.01	0.01	−0.13	0.90
Neurosciences	0.03	0.05	−0.90	0.37
Pharmacology	0.01	0.01	−0.13	0.90
Virology	0.04	0.03	0.64	0.52
Other	0.43	0.46	−0.74	0.46

Note: For the variable "Conception distinct", 180 retracted cases and 319 non-retracted cases are used for *t*-test because of missing data.

5.2. Division of labor, institutions and retractions

Table 3 shows the logistic regressions of retractions on division of labor, corruption and incentive systems as well as our control variables. First, column 1 shows that division of labor has a significant positive relation with retractions, suggesting that specialization, whether by making scientific projects vulnerable to mistakes or intentional misrepresentation, is associated with higher rates of retraction.

Columns 2 and 3 show that each institutional predictor has a significant positive relation with the rate of retraction. High corruption environments are associated with greater rates of pathologies in science, and reward policies for publication may generate pathologies as a side effect (although if they are both in the model, only corruption is statistically significant, as shown in Column 1). However, even considering these strong relations with institutional context, the division of labor still has a significant relation with retraction, independent of institutional predictors such as corruption and incentive systems. We also report a linear probability model version of our main regression (Column 6). Teams with every task performed by one member have a 22 percentage points higher probability of retraction than teams with zero tasks performed by one member (i.e., every task performed by at least two members) holding other variables constant. We also see that a one standard deviation (.92) increase in the corruption index is associated with an 8 percentage point (.92*.09) increase in the rate of retractions, while the presence of incentives (controlling for corruption, etc.) is associated with a 2 percentage points higher rate of retractions (although not statistically significant). Column 7 adds a quadratic term for division of labor. We find that there is a significant positive coefficient for the squared term, suggesting that the relation between division of labor and retractions is stronger as division of labor increases

Table 3
Regressions of retractions on division of labor, corruption and incentives.

	Logit										LPM			
	(1)		(2)		(3)		(4) Excluding China		(5) Excluding industry		(6)		(7)	
Variables	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p
Division of labor	0.94 (0.46)	0.04	0.94 (0.46)	0.04	1.05 (0.45)	0.02	1.28 (0.62)	0.04	0.91 (0.47)	0.05	0.22 (0.10)	0.04	−0.14 (0.24)	0.56
DoL squared													0.59 (0.35)	0.09
World Bank Corruption Index	0.40 (0.15)	0.01	0.44 (0.12)	0.00			0.03 (0.23)	0.89	0.32 (0.16)	0.04	0.09 (0.03)	0.01	0.09 (0.03)	0.01
Country with pub. incentives	0.10 (0.27)	0.70			0.57 (0.21)	0.01	0.17 (0.29)	0.55	0.18 (0.28)	0.53	0.02 (0.06)	0.69	0.02 (0.06)	0.74
No. of authors	−0.02 (0.02)	0.48	−0.02 (0.02)	0.49	−0.01 (0.02)	0.55	−0.01 (0.03)	0.79	−0.02 (0.03)	0.35	0.00 (0.01)	0.50	0.00 (0.01)	0.47
Interdisciplinarity	0.13 (0.07)	0.05	0.13 (0.07)	0.05	0.13 (0.07)	0.05	0.18 (0.08)	0.03	0.09 (0.07)	0.18	0.03 (0.01)	0.05	0.03 (0.01)	0.05
Top quartile JIF	0.28 (0.28)	0.32	0.27 (0.28)	0.33	0.22 (0.28)	0.44	0.46 (0.36)	0.20	0.31 (0.29)	0.28	0.06 (0.06)	0.33	0.06 (0.06)	0.33
University ranking	0.06 (0.08)	0.45	0.05 (0.07)	0.48	0.04 (0.07)	0.63	0.06 (0.08)	0.48	0.02 (0.08)	0.84	0.01 (0.02)	0.48	0.01 (0.02)	0.49
Industry affiliation	−0.16 (0.42)	0.70	−0.14 (0.42)	0.73	−0.29 (0.41)	0.49	−0.32 (0.48)	0.50			−0.03 (0.09)	0.70	−0.04 (0.09)	0.69
Publication year	−0.02 (0.04)	0.69	−0.02 (0.04)	0.69	0.00 (0.04)	0.96	−0.05 (0.05)	0.25	0.00 (0.04)	0.96	0.00 (0.01)	0.71	0.00 (0.01)	0.73
Field dummies	Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Observations	544		544		544		395		512		544		544	
LR ChiSq	29.1		28.9		21.8		19.5		24.5		R2 0.05		0.06	
DF	18		17		17		18		17					
Pr > ChiSq	0.05		0.04		0.19		0.36		0.11					

All models include intercept (not displayed).

(see below). The effect of division of labor in this model becomes positive (increases pathologies) when about 24% of tasks are solo.

Since China scores relatively high on the corruption index and also provides a cash incentive system for publication, and is also one of the largest producers of scientific papers in recent years, China may account for much of the institutional effects. Therefore, we tested our main model (i.e., column 1) excluding project teams that have any Chinese member. Column 4 shows that after excluding China, the effect of the institutional predictors disappears. However, division of labor still has a significant relation to retractions.

Since there are some concerns that industry collaborations may have higher rates of pathologies (Salandra, 2018), we also tested a model excluding papers with industry collaborators. Column 5 shows the results. Again, we find a significant positive relation between division of labor and retractions.

For controls, interdisciplinarity has a positive relation with retractions. We do not find statistically significant relations for number of authors, university ranking, industry affiliation, or high JIF papers (i.e., high competition). We further test the models with alternative measures: log of team size, mean corruption scores and direct cash incentives, using the Leiden ranking rather than the Shanghai ranking, using industry funding instead of industry affiliation, and find consistent results (results available from authors).

5.3. Alternative measures of division of labor

We constructed our division of labor measure to reflect task partitioning and redundancy, that is, how likely the task is to be conducted by more than one member, considering the theoretical arguments about the links between division of labor and pathologies, at the research team (rather than individual) level. Keeping this focus, we constructed different division of labor measures and tested them with our main model (for their descriptive statistics, see Tables 1 and 2).

Column 1 in Table 4 shows the primary measure we are using in this study. Columns 2–4 show variations on this measure. We first create a binary variable capturing whether there is any division of labor or not in the team (i.e., there is at least one task done by only one member vs. none of the tasks are done by only one member), a binary variable with a cut at the 75th percentile value (i.e., more than 25% of tasks are done by only one member vs. others), and a binary variable with a cut at the 95th percentile value (i.e., more than 60% of tasks are done by only one member vs. others). When we use a 95th percentile cut, we see a significant relation between division of labor and retractions, while the former two measures show positive, but not statistically significant, relations.

These results suggest that the relation between of division of labor and retraction could be most visible at higher levels of division of labor. Given that more than half of teams in our data have no division of labor and our measure possibly underestimates the degree of division of labor (downward bias), it may be the case that the relation is strongest as the division of labor gets larger. Furthermore, we cannot ignore the possible benefits from division of labor or specialization in teams. A person who is specialized in a certain task would have deeper knowledge in that task than a person who deals with many different tasks and hence cannot spend enough time for each task. Therefore, a certain level of specialization may prevent errors in work. However, too much

Table 4
Tests of alternative measures of division of labor.

Measure of DoL	Logit															
	(1) Primary measure		(2) Yes or No		(3) 75th percentile		(4) 95th percentile		(5) -Avg. members per task		(6) # tasks		(7) Conception distinct		(8) # diff. org	
Variables	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p	β (SE)	p
Division of labor	0.94 (0.46)	0.04	0.24 (0.19)	0.22	0.25 (0.24)	0.30	1.45 (0.55)	0.01	0.10 (0.08)	0.20	0.10 (0.13)	0.44	−0.40 (0.34)	0.24	−0.10 (0.08)	0.19
World Bank Corruption Index	0.40 (0.15)	0.01	0.41 (0.15)	0.01	0.42 (0.15)	0.01	0.39 (0.15)	0.01	0.40 (0.15)	0.01	0.41 (0.15)	0.01	0.40 (0.16)	0.01	0.41 (0.15)	0.01
Country with pub. incentives	0.10 (0.27)	0.70	0.11 (0.27)	0.70	0.08 (0.27)	0.77	0.09 (0.28)	0.75	0.10 (0.27)	0.71	0.11 (0.27)	0.70	0.04 (0.28)	0.89	0.12 (0.28)	0.67
No. of authors	−0.02 (0.02)	0.48	−0.03 (0.03)	0.29	−0.03 (0.03)	0.29	−0.02 (0.02)	0.37	0.00 (0.03)	0.92	−0.04 (0.03)	0.13	−0.03 (0.03)	0.29	−0.02 (0.03)	0.59
Interdisciplinarity	0.13 (0.07)	0.05	0.13 (0.07)	0.06	0.13 (0.07)	0.06	0.13 (0.07)	0.05	0.12 (0.07)	0.06	0.12 (0.07)	0.07	0.11 (0.07)	0.13	0.12 (0.07)	0.07
Top quartile JIF	0.28 (0.28)	0.32	0.26 (0.28)	0.36	0.26 (0.28)	0.36	0.27 (0.28)	0.34	0.25 (0.28)	0.37	0.22 (0.28)	0.44	0.07 (0.30)	0.83	0.29 (0.28)	0.31
University ranking	0.06 (0.08)	0.45	0.06 (0.07)	0.44	0.06 (0.07)	0.46	0.06 (0.08)	0.45	0.05 (0.07)	0.47	0.06 (0.07)	0.45	0.02 (0.08)	0.77	0.05 (0.08)	0.52
Industry affiliation	−0.16 (0.42)	0.70	−0.14 (0.42)	0.73	−0.14 (0.42)	0.73	−0.19 (0.42)	0.66	−0.10 (0.42)	0.81	−0.15 (0.42)	0.73	−0.25 (0.43)	0.56	−0.05 (0.42)	0.91
Publication year	−0.02 (0.04)	0.69	−0.02 (0.04)	0.72	−0.02 (0.04)	0.71	−0.01 (0.04)	0.74	−0.02 (0.04)	0.72	−0.01 (0.04)	0.72	−0.01 (0.04)	0.73	−0.01 (0.04)	0.73
Field dummies	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Observations	544		544		544		544		544		544		499		544	
LR ChiSq	29.1		26.3		25.9		32.8		26.5		25.4		21.2		26.6	
DF	18		18		18		18		18		18		18		18	
Pr > ChiSq	0.05		0.09		0.10		0.02		0.09		0.11		0.27		0.09	

All models include intercept (not displayed).

specialization may lead to members losing track of the big picture of their project, losing their sense of responsibility because of being shielded from close surveillance, and also generate structural secrecy. This can start to generate the conditions that increase the likelihood of pathologies in scientific work. Consistent with our contemplation on the positive and negative sides of specialization (Table 3, Column 7), we find that the quadratic term was positive and significant (while the linear term was negative, although not significant).

If more members work on a task, it may represent higher redundancy and a higher likelihood of the task being reviewed by other people. Therefore, we also computed the average number of members per task in each team and reverse-coded it to use it as a proxy for division of labor. Column 5 shows that the measure has a positive direction, but not significant at the conventional level. When we use the count of divided tasks (based on specified tasks in the contribution statements) in column 6, which may represent task partitioning although not task redundancy, it also shows a positive direction, but not significant.

We also considered the separation of conception from execution as a possible alternative form of division of labor, although this operationalization does not follow from our theoretical model (which is about task redundancy–i.e., persons per task rather than tasks per person). Larivière et al. (2016) show that senior scientists are more likely to engage in conception while junior scientists are more likely to engage in execution (see also Shibayama et al., 2015). Prior work suggests that a division between conception and execution of the research can have adverse effects on the productivity of scientific teams. For example, Shibayama et al. find that separation of conception from execution is associated with lower productivity for basic research, though higher productivity for applied research (Shibayama et al., 2015). Merton (1973) and Hackett (1994) argue that such separation can be a source of alienation and hence pathologies in science. However, when we tested a model (column 7) measuring whether the authors involved in conception were distinct from the authors involved in conducting the experiments, providing materials and/or analyzing the data, we find no significant relation with retractions.[11] Note that in this measure, there can be multiple authors involved in either conception, execution, or both (i.e., there can be task-level redundancy even if there is separation between those performing conception versus execution). The correlation between this separation measure and our task redundancy division of labor measure is .19. Thus, when we look at task overlap within person (whether at least one of the authors is involved in both conception and execution), we do not find any effects, in contrast to the consistent positive relation between pathologies and lack of task redundancy (exactly one author for a given task).

Lastly, we used the number of different organizations to which members in the team belong as an alternative measure in column 8 and found no clear relation. While we might think of remote collaboration as another measure of division of labor, in our conceptualization, being remote does not necessarily equal division of labor. For example, if the remote researchers are doing a specialized task (outsourcing of the analysis, for example), then remote will proxy for division of labor. However, if the remote sites are each doing the same task (for example, collecting patient data from multiple hospitals) then each site's results can be checked against the results of the other sites. In other words, task redundancy, even in remote collaborations, can reduce the chances of pathologies. In our data, the correlation between the number of different organizations and our primary division of labor measure is −.19. And, as we see in column 8, the number of different organizations (i.e., remote collaboration) does not explain retractions.

Overall, alternative measures of low task redundancy division of labor generally show a positive direction, showing qualitatively consistent results, although significance varies. The effect tends to greater statistical significance (lower p values) for measures that capture high levels of division of labor (such as the continuous measure, the quadratic measure and the 95th percentile measure). However, measures that do not capture task redundancy, such as separation of conception from execution, or remote collaboration, do not show statistically significant effects. As Bunderson and Sutcliffe (2002) pointed out, it may not be strange at all to see that these different conceptualizations of division of labor generate different results. More importantly, we have to explicitly construct the measures that best match our theory (Bunderson and Sutcliffe, 2002). We argue that our primary measure represents the characteristics of division of labor highlighted in our theoretical arguments, regarding task partitioning and redundancy, better than other alternative measures.

6. Conclusions

Steen et al. (2013) argue that: "Better understanding of the underlying causes for retractions can potentially inform efforts to change the culture of science and to stem a loss of trust in science among the lay public." In this paper, we build on organizational theory to incorporate project structure into the analysis of rates of retraction. The rise of team science, and the associated division of labor, likely produces efficiency effects on science (Hackett, 1990; Smith, 1776). At the same time, our results suggest that division of labor may make scientific findings more vulnerable to pathologies, whether mistakes or deliberate malfeasance. We find that greater division of labor is associated with greater rates of retraction.

We also find that institutional context, including national levels of corruption, and the presence of productivity-based incentives, also are associated with greater rates of retraction, but the relations with division of labor are net of these institutional contexts. We also find that interdisciplinarity is associated with greater rates of retraction, but the division of labor effect is net of interdisciplinarity. These results suggest that one explanation for the increase in retractions, and the perceived increase in pathologies in science more generally, may be that it is a side effect of increasing bureaucratization of scientific work, as well as increases in high-stakes reward systems and a globalization of science to contexts where the norms of good scientific practices may be weaker.

6.1. Limitations and future research

Assuming that reported authorship correctly reflects the contributions of all contributors, bibliometrics-based collaboration studies are growing rapidly over the last decades. A related but under-researched topic is collaboration structure, i.e., how tasks are divided among different contributors (Bodenheimer, 2000; Davis and Cronin, 1993). The published statements of the division of labor provide us a valuable source of information on collaboration structures.

However, one should not over-estimate the degree to which these contribution statements in journals are unobtrusive, objective measures. Contribution statements in journals can underrepresent actual contributions, overly simplify the information, and also over-claim the contributions of team members. In particular, there may be a socially desirable response effect, in part responding to increasing emphasis by journals and professional associations on authorship guidelines (Jabbehdari and Walsh, 2017), leading to claims that all team members participated in all of the tasks. In addition, such statements often aggregate tasks into fairly gross categories (analyzed data, conducted experiments) that may in fact be finely divided in an actual collaboration. Moreover, even if we assume the statement of division of labor captures comprehensively and accurately the real collaboration structure among contributors, the coexistence of ghost contributors (those who write or otherwise contribute to the research but do not appear as authors), or guest authors (who make minimal contributions

---

[11] We get a similar result if we compare conception with conducting the experiment or providing data (i.e., where analyzing data is not included in "execution"). Results available from the author.

but are listed as author) makes the measurement situation more complicated (Bodenheimer, 2000; Jabbehdari and Walsh, 2017). These complications in the measuring of division of labor give us a conservative test of our hypotheses due to a downward bias of the division of labor measure. Even with these limitations in our measure, we found significant association between division of labor and retractions. However, future research needs a cleaner and better measure of division of labor to address these problems.

It is also the case that retractions are an imperfect measure of scientific pathologies. It is likely that many pathologies do not get observed, or if noticed, are not constructed as rising to the level of a serious enough concern to require a public statement of retraction. There may be opportunities using the tools of data analytics to gain new insights into possible markers of error or manipulations in data (Bergh et al., 2017; Park et al., 2017). Hence, we hope that future work refines both the measures of the structure of scientific work and the likelihood of pathologies in published findings. Finally, it is important to note that these data come primarily from biomedical research. While the theory is developed based on prior work in organizations spanning a wide variety of work settings (including engineering, finance, construction, medicine, etc.) and then tested on biomedical researchers, it is still an open question as to whether the results would hold in scientific fields with different work organizations. In particular, the levels of uncertainty and interdependence in different fields might affect these relationships (cf. Walsh and Lee, 2015, for a discussion of these issues). Hence, future work examining other fields of science is needed to see how generalizable these findings are.

Furthermore, in addition to the structural and contextual factors tested here, there may be aspects of the research topic itself that increase the rates of pathologies, such as novelty or complexity. We have used a variety of methods to control for these alternative potential causes, such as matching, as well as controlling for interdisciplinarity, field, journal impact factor, university rank, and publication year. However, given the observational nature of our data, we cannot rule out all potential confounding factors related to unmeasured heterogeneity across projects.

Given these limitations in measures and in methods, further work is needed to see how the specifics of the research team structure (for example, the degree of hierarchy) may affect the rates of scientific pathologies. In addition, ethnographic or historical work on pathological cases may highlight the specific processes by which division of labor may result in scientific pathologies (or, put differently, how more task integration may help detect mistakes or even deliberate malfeasance).

6.2. Policy implications and final remarks

There has been increasing emphasis in the science policy community on training in research ethics. For example, NIH and NSF now require Responsible Conduct of Research training for graduate students and principle investigators. However, much of this training involves an individual-centered approach, emphasizing the need to be ethical in one's own conduct (National Academy of Sciences, 2009). Similarly, a recent discussion of pathologies in management research proposed two new policy recommendations that emphasized a combination of sensitivity to the problem (an author checklist) and increased enforcement (routine use of plagiarism detection software) (Harley et al., 2014). However, we need to move beyond individual-level interventions. The organization theory of scientific pathologies and the empirical support

consistent with the theory suggest the need for greater emphasis on the shared responsibility for team research findings (Chubin, 1985). For example, Chubin reports the following suggestions based on an NIH review of a high profile cases of scientific pathologies: having multiple members of the team conduct the experiments; having more interactions between PIs and other members of the lab; sharing lab notebooks among team members; and posting explicit guidelines about research conduct. While these suggestions were proposed decades ago, such suggestions have yet to be incorporated as best practice. It is still the case that current NAS guidelines emphasize individual responsibility and university culture rather than changing the structure of projects to reduce the vulnerability to pathologies (NAS, 2017). There is some suggestion in the recent National Academies report on research integrity noting that "collaborative research must include structures that coordinate and verify the integrity of separate contributions to the overall research effort. (p.16)" (NAS, 2017). The NAS report suggests that scientific collaborations would benefit from incorporating practices from high-reliability systems (Paté-Cornell, 1990; Reason, 2000). A report from the US Geological Survey on a notorious case of failure to detect longstanding data fabrication in one of its testing labs notes the importance of these quality control methods for detecting and preventing errors from testing labs (whether these are deliberate or accidental) (Department of the Interior OIG, 2016). Cross-training, rechecking, and duplicate parallel production are all quality control practices that might reduce the chances of pathologies, but each of these is costly and time consuming and may be sacrificed in favor of a culture of trust, and an emphasis on productivity. Hence, science funders might consider changing expectations for production to emphasize the checks and balances within projects that might detect pathologies before they become published, even if such changes increase the cost and decrease the productivity of scientific projects. Given the increase in team science, there is a growing need for best practice guidelines for the responsible and effective conduct of *team* research. Similarly, as Larkin and Pierce (2016) find, reward structures that incentivize productivity (and the behaviors that will increase productivity) are also likely to incentivize pathologies. Our findings here suggest that division of labor as a solution to the demand for increasing productivity might also be pathogenic. Hence, when designing incentive systems in science, we should be aware of their potential for inducing pathological outcomes.

Following recent work in the organization theory of science and of pathologies in science, we have made an initial attempt to explore empirically the relationship between collaboration structure (based on retrieved division of labor information) and the likelihood of scientific pathologies. We hope this will open new avenues for research in science, and in other technical settings, that use the structure of the work to explain pathologies in terms of the organization of the work rather than simply individual wrongdoings. This more sociological perspective on pathologies may provide greater insights into when such pathologies are more or less likely to occur. It can also guide managers and policy makers toward additional means for reducing pathologies.

Acknowledgements

Appendix A

Table A1
Examples of different wordings for tasks.

Tasks	Similar wordings
Conceive	Conceived of the study; designed the study/the project/research/experiments; study initiation; contributed to experimental design; contributed to the conception; planned the study; developed the trial protocol
Perform	Acquired the data; involved in data collection; performed the experiment; involved in clinical aspects; conducted fieldwork; completed all image processing for the study
Provide materials/data	Contributed new reagents/materials/analytic tools; provided bacterial strains; carried out the arrangement of the enrolled patients; contributed clinical samples; provided the metabolomics technology; provided technical support with respect to the electron microscopy
Analyze	Analyzed data; performed the statistical analysis; interpreted results of experiments; prepared figures; western blot assays;
Write	Wrote the paper; drafted the manuscript; prepared the manuscript; responsible for revisions of draft manuscripts; reviewed the article; critically reviewed the manuscript drafts
Other	Financial support; administrative support; provided supervision

Notes:

1. Some contribution statements say, e.g., "These authors contributed equally". This statement is related to authorship, but does not provide any information on task distribution among authors. Therefore, we treat these cases as missing in the division of labor measure.

2. The usual declaimer such as "All authors were responsible for approval of the final manuscript." is ignored because it does not provide any specific information (See Table B1 in Appendix B, Case D).

3. We define "carried out the arrangement of the enrolled patient" as "provide materials" because in biomedical research this task is mostly done by a private provider, which is a very separate job, and the enrolled patient is materials for biomedical research.

4. When a member "provided supervision," and is not included in any other task but supervision or only serving as a lab head, it is classified into "Other".

5. Five percent include "Other" as one of the divided tasks.

Appendix B

Table B1
Construction of a division of labor measure.

	Contribution statements	Conceive	Perform	Materials, data, sw	Analyze	Write	Other
Case A	P.S. conceived and designed the study, provided financial support, assembled, analyzed, and interpreted the data, and wrote the paper; E.M., M.G.d.I., E.R., and F.C. carried out the experiments, and collected, assembled, analyzed, and interpreted data; M.T. and V.G. were responsible for provision of study material, and analyzed and interpreted data; and G.Z. conceived and designed the study, provided financial support, wrote the paper, and was responsible for final approval of the paper.	PS, GZ	PS, EM, MGDI, ER, FC	MT, VG	PS, EM, MGDI, ER, FC, MT, VG	PS, GZ	PS, GZ
Case B	Y.-W.Q. conceptualized and designed the research, designed and performed the experiments, analyzed and interpreted the results, and wrote the paper; Y.C., B.H., and Y.Z. performed experiments; B.S. aided experimental design; B.B. contributed clinical samples and wrote the paper; J.D.S. conceptualized and designed the research, analyzed and interpreted the results, and wrote the paper.	YWQ, BS, JDS	YWQ, YC, BH, YZ	BB	YWQ, JDS	YWQ, BB, JDS	
Case C	XBX collected material, carried out statistical analysis and wrote the manuscript; HZ provided evaluating data; PZ help to complete the article; LC was involved in editing the manuscript and statistical analysis; all authors provided final approval to the submitted version of the article.		XBX	HZ	XBX, LC	XBX, LC, PZ	
Case D	XHJ designed research; BQY performed research; WQ contributed analytic tool and analyzed data; JWH wrote the paper. All authors read and approved the final manuscript.	XHJ	BQY	WQ	WQ	JWH	

Case A: $DoL = 1 - \frac{\text{\# of tasks perfomed by at least two members}}{\text{Total \# of tasks presented in contribution}} = 1 - \frac{6}{6} = 0$.

Case B: $DoL = 1 - \frac{\text{\# of tasks perfomed by at least two members}}{\text{Total \# of tasks presented in contribution}} = 1 - \frac{4}{5} = 0.2$.

Case C: $DoL = 1 - \frac{\text{\# of tasks perfomed by at least two members}}{\text{Total \# of tasks presented in contribution}} = 1 - \frac{2}{4} = 0.5$.

Case D: $DoL = 1 - \frac{\text{\# of tasks perfomed by at least two members}}{\text{Total \# of tasks presented in contribution}} = 1 - \frac{0}{5} = 1$.

For example, for Case B, 20% of divided tasks are performed by only one member.

References

Azoulay, P., Bonatti, A., Krieger, J.L., 2015a. The Career Effects of Scandal: Evidence from Scientific Retractions. National Bureau of Economic Research Working Paper #21146.

Azoulay, P., Furman, J.L., Krieger, J.L., Murray, F., 2015b. Retractions. Rev. Econ. Stat. 97, 1118–1136.

Baker, W.E., Faulkner, R.R., 1993. The social organization of conspiracy: illegal networks in the heavy electrical equipment industry. Am. Sociol. Rev. 58, 837–860.

Baltimore, D., 1991. Dr Baltimore says "sorry". Nature 351, 94–95.

Baron, J.N., King, M., Sorenson, O., 2016. S/he blinded me with science. In: Palmer, D., Greenwood, R., Smith-Crowe, K. (Eds.), The Sociology of Scientific Misconduct. Organizational Wrongdoing Cambridge University Press.

Beasley, M.R., Datta, S., Kogelnik, H., Kroemer, H., Monroe, D., 2002. Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors. Lucent Technologies.

Becker, H.S., 1963. Outsiders. Free Press, New York.

Belot, M., Schröder, M., 2013. Sloppy work, lies and theft: a novel experimental design to study counterproductive behaviour. J. Econ. Behav. Organiz. 93, 233–238.

Bergh, D.D., Sharp, B.M., Li, M., 2017. Tests for identifying "Red flags" in empirical findings: demonstration and recommendations for authors, reviewers, and editors. Acad. Manage. Learn. Educ. 16, 110–124.

Bhattacharjee, Y., 2013. The Mind of a Con Man. New York Times, New York.

Biagioli, M., 2003. Rights or rewards? Changing frameworks of scientific authorship. In: Biagioli, M.G., Peter (Eds.), Scientific Authorship: Credit and Intellectual Property in Science. Routledge, New York, pp. 253–280.

Blau, P.M., 1970. A formal theory of differentiation in organizations. Am. Sociol. Rev. 35, 201–218.

Bodenheimer, T., 2000. Clinical investigators and the pharmaceutical industry. N. Engl. J. Med. 343, 511.

Brass, D.J., Butterfield, K.D., Skaggs, B.C., 1998. Relationships and unethical behavior: a social network perspective. Acad. Manage. Rev. 23, 14–31.

Braverman, H., 1974. Labor and Monopoly Capital. Monthly Review Press, New York.

Bunderson, J.S., Sutcliffe, K.M., 2002. Comparing alternative conceptualizations of functional diversity in management teams: process and performance effects. Acad. Manage. J. 45, 875–893.

Butler, N., Delaney, H., Spoelstra, S., 2017. The gray zone: questionable research practices in the business school. Acad. Manage. Learn. Educ. 16, 94–109.

Carey, B., Belluck, P., 2015. Doubts About Study of Gay Canvassers Rattle the Field. The New York Times, New York.

Chinoy, E., 1955. Automobile Workers and the American Dream. Doubleday & Company, Garden City, NY.

Chubin, D.E., 1985. Research malpractice. BioScience 35, 80–89.

Chubin, D.E., Hackett, E.J., 1990. Peerless Science. State University of New York Press, Albany, NY.

Corrêa Jr., E.A., Silva, F.N., da, F., Costa, L., Amancio, D.R., 2017. Patterns of authors contribution in scientific manuscripts. J. Informet. 11, 498–510.

Crane, D., 1972. Invisible Colleges. University of Chicago Press, Chicago.

Davis, C.H., Cronin, B., 1993. Brief communication acknowledgments and intellectual indebtedness: a bibliometric conjecture. J. Am. Soc. Inf. Sci. 44 (1986-1998), 590.

Department of the Interior OIG, 2016. Inspection of Scientific Integrity Incident at USGS Energy Geochemistry Laboratory.

Durkheim, E., 1951. Suicide. Free Press, New York.

Fanelli, D., 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One 4, e5738.

Fang, F.C., Steen, R.G., Casadevall, A., 2012. Misconduct accounts for the majority of retracted scientific publications. Proc. Natl. Acad. Sci. 109, 17028–17033.

Fisher, J.A., 2009. Medical Research for Hire. Rutgers University Press, New Brunswick, NJ.

Fox, M.F., Braxton, J.M., 1994. Misconduct and social control in science: issues, problems, solutions. J. High. Educ. 65, 373–383.

Franzoni, C., Scellato, G., Stephan, P., 2011. Changing incentives to publish. Science 333, 702–703.

Furman, J.L., Jensen, K., Murray, F., 2012. Governing knowledge in the scientific community: exploring the role of retractions in biomedicine. Res. Policy 41, 276–290.

Goodman, P.S., Ramanujam, R., Carroll, J.S., Edmondson, A.C., Hofmann, D.A., Sutcliffe, K.M., 2011. Organizational errors: directions for future research. Res. Organiz. Behav. 31, 151–176.

Greve, H.R., Palmer, D., Pozner, J., 2010. Organizations gone wild: the causes, processes, and consequences of organizational misconduct. Acad. Manage. Ann. 4, 53–107.

Hackett, E.J., 1990. Science as a vocation in the 1990s: the changing organizational culture of academic science. J. High. Educ. 241–279.

Hackett, E.J., 1994. A social control perspective on scientific misconduct. J. High. Educ. 65, 242–260.

Haeussler, C., Sauermann, H., 2016. The Division of Labor in Teams: A Conceptual Framework and Application to Collaborations in Science. National Bureau of Economic Research.

Halle, D., 1984. American'S Working Man. University of Chicago Press, Chicago.

Harley, B., Faems, D., Corbett, A., 2014. A few bad apples or the tip of an iceberg? Academic misconduct in publishing. J. Manage. Stud. 51, 1361–1363.

Honig, B., Bedi, A., 2012. The fox in the hen House: a critical examination of plagiarism among members of the academy of management. Acad. Manage. Learn. Educ. 11, 101–123.

Honig, B., Lampel, J., Siegel, D., Drnevich, P., 2014. Ethics in the production and dissemination of management research: institutional failure or individual fallibility? J.

Manage. Stud. 51, 118–142.

Hughes, E.C., 1984. Mistakes at Work, The Sociological Eye. Transaction Books, New Brunswick, NJ, pp. 316–325.

Jabbehdari, S., Walsh, J.P., 2017. Authorship norms and project structures in science. Sci. Technol. Hum. Values, 0162243917697192.

Jin, G.Z., Jones, B., Lu, S.F., Uzzi, B., 2013. The Reverse Matthew Effect: Catastrophe and Consequence in Scientific Teams, NBER Working Paper.

Jones, B.F., 2009. the burden of knowledge and the death of the renaissance Man: Is innovation getting harder? Rev. Econ. Stud. 76, 283–317.

Lacetera, N., Zirulia, L., 2011. The economics of scientific misconduct. J. Law Econ. Organiz. 27, 568–603.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., Sugimoto, C.R., 2016. Contributorship and division of labor in knowledge production. Soc. Stud. Sci. 46, 417–435.

Larkin, I., Pierce, L., 2016. Compensation and employee misconduct. In: Palmer, D., Smith-Crowe, K., Greenwood, R. (Eds.), The Inseparability of Productive and Counterproductive Behavior in Firms. Organizational Wrongdoing. Cambridge University Press, Cambridge, UK, pp. 270–304.

Latour, B., Woolgar, S., 1979. Laboratory Life. Sage, Beverly Hills.

Lee, Y.-N., Walsh, J.P., Wang, J., 2015. Creativity in scientific teams: unpacking novelty and impact. Res. Policy 44, 684–697.

Lerner, A.C., 2013. Statement of Allison C. Lerner, Inspector General, National Science Foundation, Committee on Science, Space, and Technology Subcommittee on Investigations and Oversight United States House of Representatives: Top Challenges for Science Agencies,Reports from the Inspectors General.

Lewellyn, K.B., Judge, W.Q., Smith, A., 2017. Exploring the questionable academic practice of conference paper double dipping. Acad. Manage. Learn. Educ. 16, 217–236.

Lewis, J., 2015. Research policy as "carrots and sticks": governance strategies in Australia, the United Kingdom and New Zealand. In: Capano, G., howlett, M., Ramsh, M. (Eds.), Varieties of Governance: Dynamics, Strategies, Capacities. Palgrave, Hampshire, UK, pp. 131–150.

Liu, W., Hu, G., Gu, M., 2016. The probability of publishing in first-quartile journals. Scientometrics 106, 1273–1276.

Lu, S.F., Jin, G.Z., Uzzi, B., Jones, B.F., 2013. The retraction penalty: evidence from the web of science. Sci. Rep. 3https://doi.org/10.1038/srep03146. 3146.

Macaluso, B., Larivière, V., Sugimoto, T., Sugimoto, C.R., 2016. Is science built on the shoulders of women? A study of gender differences in contributorship. Acad. Med. 91, 1136–1142.

Merton, R.K., 1938. Social structure and anomie. Am. Sociol. Rev. 3, 672–682.

Merton, R.K., 1949. Social Theory and Social Structure. Free Press, Glencoe, IL.

Merton, R.K., 1973. The Sociology of Science. The University of Chicago Press, Chicago.

Meyer, J.W., Rowan, B., 1977. Institutionalized organizations: formal structure as myth and ceremony. Am. J. Sociol. 83, 340–363.

Milojević, S., 2014. Principles of scientific research team formation and evolution. Proc. Natl. Acad. Sci. 111, 3984–3989.

Mohliver, A., 2018. How misconduct spreads: auditors' role in the diffusion of stock-option backdating. Adm. Sci. Q., 0001839218763595.

Mongeon, P., Larivière, V., 2016. Costly collaborations: the impact of scientific fraud on co-authors' careers. J. Assoc. For. Inform. Sci. Technol. 67, 535–542.

Morgan, S.L., Winship, C., 2012. Counterfactuals and Causal Inference. Cambridge University Press.

Nagaoka, S., Owan, H., 2014. Author Ordering in Scientific Research: Evidence from Scientists Survey in the US and Japan," IIR Working Paper. Institute of Innovation Research, Hitotsubashi University, Tokyo, Japan.

National Academy of Sciences, 2009. On Being a Scientist: a Guide to Responsible Conduct in Research, third edition. National Academies Press, Washington, DC.

National Academy of Sciences, Ea.M., 2017. Fostering Integrity in Research. The National Academies Press, Washington, DC.

Necker, S., 2014. Scientific misbehavior in economics. Res. Policy 43, 1747–1759.

Noyori, R., Richmond, J.P., 2013. Ethical conduct in chemical research and publishing. Adv. Synth. Catal. 355, 3–9.

Osterloh, M., Frey, B.S., 2015. Ranking games. Eval. Rev. 39, 102–129.

Park, J., Howe, J.D., Sholl, D.S., 2017. How reproducible are isotherm measurements in metal–organic frameworks? Chem. Mater. 29, 10487–10495.

Paté-Cornell, M.E., 1990. Organizational aspects of engineering system safety: the case of offshore platforms. Science 250, 1210–1217.

Perrow, C., 1984. Normal Accidents: Living With High Risk Systems. Basic Books, New York.

Porter, A., Rafols, I., 2009. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. Scientometrics 81, 719–745.

Quan, W., Chen, B., Shu, F., 2017. Publish or Impoverish: An Investigation of the Monetary Reward System of Science in China (1999–2016). DOI: arXiv:1707.01162.

Reason, J., 2000. Human error: models and management. BMJ: Br. Med. J. 320, 768.

Riemer, J.W., 1976. "Mistakes at work" the social organization of error in building construction work. Soc. Problems 23, 255–267.

Roberts, K.H., Libuser, C., 1993. From Bhopal to banking: organizational design can mitigate risk. Organiz. Dyn. 21, 15–26.

Robinson, S.L., Bennett, R.J., 1995. a typology of deviant workplace behaviors: a multi-dimensional scaling study. Acad. Manage. J. 38, 555–572.

Roth, J.A., 1966. Hired hand research. Am. Sociol. 1, 190–196.

Salandra, R., 2018. Knowledge dissemination in clinical trials: exploring influences of institutional support and type of innovation on selective reporting. Res. Policy 47 (7), 1215–1228.

Shibayama, S., Baba, Y., Walsh, J.P., 2015. Organizational design of university laboratories: task allocation and lab performance in Japanese bioscience laboratories. Res.

Policy 44, 610–622.

Sismondo, S., 2009. Ghosts in the machine: publication planning in the medical sciences. Soc. Stud. Sci. 39, 171–198.

Smith, A., 1776. The Wealth of Nations.

Steen, R.G., 2010. Retractions in the scientific literature: is the incidence of research fraud increasing? J. Med. Ethics 2010.040923.

Steen, R.G., Casadevall, A., Fang, F.C., 2013. Why has the number of scientific retractions increased? PloS One 8, e68397.

Tang, L., Shapira, P., Youtie, J., 2015. Is there a clubbing effect underlying Chinese research citation increases? J. Assoc. For. Inf. Sci. Technol. 66, 1923–1932.

Tang, L., Hu, G., Sui, Y., Cao, C., 2016. Retraction: The Other Face of Research Collaboration, Working Paper.

Vaughan, D., 1999. The dark Side of organizations: mistake, misconduct, and disaster. Annu. Rev. Sociol. 25, 271–305.

Wagner, C.S., Leydesdorff, L., 2005. Network structure, self-organization, and the growth of international collaboration in science. Res. Policy 34, 1608–1618.

Walsh, J.P., Lee, Y.-N., 2015. The bureaucratization of science. Res. Policy 44, 1584–1600.

Warren, D.E., 2003. Constructive and destructive deviance in organizations. Acad. Manage. Rev. 28, 622–632.

Whitley, R., 2007. Changing governance of the public sciences. In: Whitley, R., Gläser, J. (Eds.), The Changing Governance of the Sciences: The Advent of REsearch Evaluation Systems. Springer, Netherlands, pp. 1–25.

World Bank, 2016. Description of Methodology–Control of Corruption Index. Worldwide Governance Indicators. Retrieved from:. http://info.worldbank.org/governance/wgi/-home.

Wuchty, S., Jones, B., Uzzi, B., 2007. The increasing dominance of teams in the production of knowledge. Science 316, 1036–1039.

Zuboff, S., 1988. In the Age of the Smart Machine: The Future of Work and Power. Basic, New York. .

Zuckerman, H., 1988. The sociology of science. In: Smelser, N.J. (Ed.), Handbook of Sociology. Sage, Newbury Park, CA, pp. 511–574.