Title
Fraud by Numbers: Metrics and the New Academic Misconduct

Author
Biagioli, Mario

Publication Date
2020-09-07

Peer reviewed

Fraud by Numbers: Metrics and the New Academic Misconduct

September 7, 2020 • By Mario Biagioli

Banner image: Man with a Beard, *generally accepted to be a Rembrandt painting until the 1940s.*

¤

Gaming excellence

According to Goodhart's Law, as soon as a measure becomes a target, gaming ensues, which undermines its function as a measure. Charles Goodhart, an economist, was referring to the gaming of economic indicators, but his law applies equally well to all sorts of regimes of evaluation, including the metrics that command so much authority in today's higher education. Universities are investing ever more heavily in curating and occasionally faking figures that enhance their national and global rankings, while simultaneously keeping those metrics in mind when deciding anything from campus development projects to class size. [1] (Architecturally

ambitious campuses attract alumni giving, which is a positive factor in the *U.S. News & World Report* rankings of universities, as are classes capped at 19 students.) Now in full swing, this trend started inconspicuously a few decades ago. Already in 1996, Northeastern University's president, Richard Freeland, observed that "schools ranked highly received increased visibility and prestige, stronger applicants, more alumni giving, and, most important, greater revenue potential. A low rank left a university scrambling for money. This single list […] had the power to make or break a school." [2] Freeland quickly figured out which numbers Northeastern needed to privilege. Ranked 162nd in 1996, Northeastern jumped to 98th in 2006 and, ten years after his departure, 47th in 2016.

This trend goes hand in hand with another distinctive feature of the modern university: the discourse of excellence. Because "excellence" is devoid of a referent that can be either empirically or conceptually defined — its meaning effectively boiling down to "being great at whatever one may be doing" — it can only be conjured into existence through quantitative indicators of more or less related proxies such as faculty productivity, their publications' impact, students' admission rates and average SAT scores, the university's expenditure per student, alumni donations, numbers of books in the library, grant funding secured by the faculty, students' post-graduation employment rate, faculty salary, and so on. The semantic pliability of excellence makes it like warm wax. Having no inherent form of its own, it can only assume the shapes that different metrics stamp on it. This symbiosis explains the ubiquity of metrics and rankings in the so-called university of excellence. [3] (The concept of "impact" shares the same definitional and semantic slipperiness as "excellence"; it depends on metrics for its existence, if only as a number in search of meaning. [4])

Responses to the modern university's embrace of "audit culture" — the shift from traditional qualitative forms of evaluation to quantitative indicators of excellence or impact — have been

polarized, especially when metrics are deployed to evaluate specific publications or individual authors through the citations they receive. [5] Advocates praise the transparency and accountability that metrics are claimed to provide, which they contrast with the traditional, opaque, and discipline-specific protocols of peer review. Critics see metrics instead as the predictable outcome of the contemporary university's managerial culture. Hiding behind the rhetorical shield of objectivity, metrics function, these critics argue, as disciplinary techniques while failing to measure anything worth measuring. What has gone unnoticed in these debates, however, is that regardless of whether citation-based indexes are a paragon of objectivity or mere numerical simulacra of imaginary value, they have an uncanny ability to spawn new and expanding forms of research misconduct. It's worth carefully analyzing such practices to understand the ways in which concepts like value and quality, and "publication" are being radically reshaped in the age of metrics.

Postproduction misconduct

The assumption has been that research misconduct emerges in response to "publish or perish" environments, and that it is epistemic in nature. The US federal government defines research misconduct as fabrication, falsification, and plagiarism in the production and publication of research results — three forms of manipulation that result in false representations of evidence, claims, and authorship. This traditional kind of research misconduct continues seemingly unabated, but we are now witnessing the emergence of new forms of manipulation — postproduction misconduct — that do not fake the content of a publication but rather enhance its impact or metric footprint. [6]

Fake peer reviews are an example. When journal editors ask you to provide the names of possible reviewers for your manuscript (which happens surprisingly often, due to many editors' limited knowledge of specialized subfields), you might decide to offer the names of real

scientists but provide non-institutional email addresses for them, like john.doe@gmail.com. That account does not actually belong to Dr. Doe — you've opened it under his name in order to receive a request to review your own manuscript. At least some of the reports are now predictably positive, allowing your article to secure publication in a higher impact journal than it otherwise might have, thus generating extra visibility and citations for that publication. Another example are citation rings, where you and other scholars agree to cite each other's articles extensively, even when the topics are unrelated. [7] (Some of these practices emerge in response to the establishment of citation targets by national governments and funding agencies. [8]) And then there are coercive citations, quid pro quos involving peer reviewers or even editors who encourage "the authors of work under review to cite the reviewers' own research in exchange for positive reviews." [9] (These requests do not just involve a handful of citations but an average of 35 per paper.)

You might also decide to make up collaborators out of thin air — fake co-authors who contribute to your article's impact by being falsely associated with prestigious universities. (In one case, a scientist made up three co-authors — "D. Wilson," "W. Wang," and "P. L. Richardson" — endowing them with faculty positions at Caltech.) The presence of such fake collaborators facilitates the publication and visibility of your article thanks to the institutional brands they are associated with. [10]

A more radical approach involves hacking journal databases, finding articles already reviewed and accepted, and adding your name to the existing authors' byline right before the article goes to press. Unless the article has many co-authors, you will surely be found out and the article may eventually be retracted, but that will take at least a few years. In the meantime, you have one extra publication and a substantial head start. You have invested almost no time in that article (much less time than it would have taken you to write and publish a paper, even if you

used fraudulent evidence to cut down on laboratory time), and yet you will start to accrue citations in a month or so when the piece is published.

Finally, if you are more risk-averse and have some disposable cash, you can simply buy a place on the author byline of articles already written and submitted for publication by writing companies. These companies exploit the fact that journals sometimes allow changes in the byline while the manuscript is being revised — they can thus sell a few extra co-authorship slots, indexing the price to the ranking of the journal. [11] You are literally paying for the citations you are expected to get.

Journals and universities are not always above these practices. Editors increase their journal's impact factor by entering into co-citation agreements with other journal editors, which involves having the authors of one journal cite articles from the colluding journal; some universities invite their faculty to cite each other to help improve the institution's rankings; and other universities pay productive scholars from other institutions generous salaries for short visits that are in fact a pretext for listing their publications and citation counts as if they were members of the permanent faculty. (More proactive researchers do not wait to be offered such deals but effectively auction off their publication list to the highest bidder, which may turn out not to be their home institution. [12]) Either way, these strategies have allowed for remarkably fast climbs, like that of Jeddah's little-known King Abdulaziz University, which managed to surpass Cambridge University in the 2014 *U.S. News & World Report*'s ranking for graduate programs in mathematics. [13]

These cases, and those mentioned earlier, have elicited outright condemnation and some cynical laughter at the creativity involved, but such responses deflect attention away from what is most worrisome — namely, how easily these metrics can be gamed or made up, in so many different ways. Postproduction manipulations exemplify the proverbial moving target, rapidly mutating

to exploit loopholes in existing indicators or in new ones produced by competition among rating providers. Once you realize the variety of metrics-induced manipulations, the copycat patterns that both modify and disseminate them, and the fact that attempts to prevent gaming only create opportunities for new manipulations, it becomes exceedingly difficult to believe that the cat could ever be put back in the bag. As economist W. Brian Arthur reminds us, "All systems will be gamed." [14] There will be many systems to be gamed in the increasingly complex regime of academic metrics, and the ever-expanding repertoire of fraudulent moves and techniques are likely to make the old triad of traditional fraud — fabrication, falsification, and plagiarism — look retro, almost quaint.

Financializing impact?

Manipulations of metrics are unethical or worse and yet do not fit the conceptual template of the US federal definition of research misconduct because they do not involve the falsification or fabrication of claims and evidence. The new misconduct, in fact, resembles the manipulation of Facebook or Instagram "likes" or other online ratings more than it does traditional research fraud and misconduct.

Though it misrepresents the authorship of the publication, adding fake co-authors does not amount to plagiarism because no appropriation is involved. (You are not lifting ideas or text from anyone but rather sharing your authorial credit with persons who do not in fact exist, which means that you are effectively neither taking nor giving anything.) And while making up fake co-authors affiliated with real prestigious institutions may infringe the trademarks of those universities, such harm falls well outside of the definition of research misconduct. Similarly, adding your name to an article authored by others involves neither falsification nor fabrication of evidence or claims (you touch nothing except the authors' byline), nor does it amount to traditional plagiarism because, while you are fraudulently claiming co-authorship (not to

mention breaking into a proprietary database), you have not technically stolen text or ideas. To make an avian analogy, you are not stealing an egg from another bird's nest, but rather putting your egg in theirs. More than a thief or a fraud, you are a parasite.

We do not know the precise motives behind all of these practices, though some, like implanting or purchasing authorship in extremis, seem correlated to urgent professional deadlines involving metrics — the number of articles one is expected to publish within a certain period and the ranking of the journals where those articles should be published. There are also various indicators of the longer-term impact of one's entire body of work like the h-index, Google Scholar's i10-index, and others, but no matter the specific index or deadline, the publication game boils down to beefing up virtual indicators of impact or, more importantly, conducting transactions based on them. "Scientometrics" was originally conceived as a literature-indexing and search technique, quickly evolved into a set of evaluative tools, and has now become the infrastructure upon which a new full-fledged economy of academic value and credit is being articulated.

For example, having a number of articles in top high-impact journals is likely to translate not only into a good tenure-track position, but also into a promotion, salary increase, or, in some universities, a cash bonus. This is costly for the university, but the researcher's high-impact publications will contribute to the university's ranking, and thus to its ability to attract donors, contracts, and top (and in some cases top-paying) students who, more than ever, rely on metrics and rankings to orient their college applications.

Citations have now become tokens of value or impact that, like other data in the age of its monetization, can be effectively bought and sold, turned into indexes, repackaged and resold to new users for uses far removed from the evaluation of a specific publication or author. There is, of course, no guarantee that promises of impact will materialize exactly as predicted, but the

academic metrics system has already developed instruments that enable transactions to happen under conditions of uncertainty. These instruments effectively separate the object of the transaction from the assets — the actual citations. Both praised and hated, the journal impact factor (JIF) and its younger sibling (Elsevier's CiteScore) are the best examples of such instruments.

Defined as the total number of times a journal's articles were cited over a two-year period divided by the number of citable items published in that journal over the same period, the impact factor effectively signifies a journal's citation "density." But while the JIF is calculated on the citations received by previously published articles (thus functioning as an index of the journal's past impact), in practice it is attached to individual articles that have just been published in that journal — articles that have not yet received the number of citations they are expected to receive based on the journal's historical averages. Despite its name, the JIF is routinely used "off label" as an article's impact factor — an impact conferred on that article at birth, before the article has any measurable impact whatsoever.

The unofficial, de facto reframing of the JIF as a rating that can be transferred to articles entails disregarding the paradox of attributing impact to something that, according to the very definition of impact, cannot yet have any. That bit of magic, however, goes a long way toward explaining the JIF's success. Conferring on the article the impact factor of the journal shaves about five to eight years off the slow process of impact and value accumulation. That is a game changer, jump-starting career opportunities and rewards. No wonder junior researchers support the journal impact factor. [15]

But while the JIF can accelerate time for scientists and their institutions, it cannot do the same for the journals themselves. No matter their quality, new journals are penalized by their short track record, which translates into very low JIFs. Potentially impactful submissions are rerouted

toward more established venues with higher impact factors, creating a negative feedback loop for the new entrants. That their JIF will improve rather slowly over several years or decades may explain some of the co-citation cartels between journals as a way to speed up the process. An even faster acceleration may be achieved by purchasing a desirable impact factor from a variety of less-than-reputable rating services that work very hard at passing themselves off as the real ones. [16]

As important as its time-accelerating effects is the fact that the JIF, in addition to having been made to price articles rather than just journals, has become the standard currency in which that price can be expressed. It is the "dollar" of the new economy of academic metrics and, whether they like its effects or not, everybody — authors, journals, universities, funding agencies — knows that. One does not even need to understand how the JIF is calculated or how reliable it is because most academic institutions will accept it as "legal tender" of impact. The JIF, therefore, does not simply rank journals but helps to create the conditions of possibility for a market and transactions based on it.

For instance, according to the journal impact factor, your *Nature* article was expected to gain about 43 citations over two years, but it received only a handful. That may be a problem for *Nature* if it keeps publishing many similarly underperforming articles, eventually causing its impact factor to dip. But it is not an issue for the author who, by the time the actual citations trickle in and convey a finer-grained sense of the article's impact, may have already received a job, tenure, or an award based, in part, on the JIF associated with that publication. And it probably did not hurt the university either. One can imagine the chair of the department sending out an annual fundraising letter to donors, proudly listing how many *Nature* articles (treated as tokens of value, according to their JIF) the faculty published that year. How many citations those papers have actually received does not matter in that context.

The reason why the faculty, the university, and the donors can do academic or philanthropic business with each other without worrying about the exact number of citations these articles have received or will receive is because those transactions are based on the impact factor of the journal where the articles were published, not their actual citation counts. The JIF is by no means independent of the articles' actual citations — it is in fact produced by them — but, importantly, it is also insulated from fluctuations in the citational performance of specific articles. As with the Black-Scholes-Merton formula for pricing options whose introduction and impact on financial markets have been studied by sociologist Donald MacKenzie, if all concerned agree that the journal impact factor sets the value — the price, in effect — of an article's impact, the market will come to perform according to the JIF, which it seems indeed to have done. [17]

Similarly, we have learned that the cost of purchasing a slot on the author byline is often indexed to the journal impact factor, and a growing number of universities now give their faculty (and occasionally their students) cash bonuses indexed on the JIF — up to $43,000 for a *Science* or *Nature* paper, but with reports of occasional six-digit awards. [18] (Cash bonuses are highest in China, followed by Saudi Arabia, Qatar, Malaysia, Taiwan, the United Kingdom, and the United States, clearly evidencing the correlation between universities' determination to quickly climb in the global rankings and their willingness to open their purse to get the metrics they need to fuel the climb. [19]) So, if the pricing is done right, the scientist who paid $5,000 for the co-authorship of an article in a 15-JIF journal may receive a comparable cash bonus from his university for having published an article in that journal.

We do not yet have evidence of arbitrage involving journal impact factors — like finding a writing company that sells you authorship at a price lower than the bonus your university will give you for publishing that same article. (Sadly, given the entrepreneurial spirit at work in

metrics-related gaming and misconduct, that would not be beyond the realm of possibility.) Interestingly, were such an arbitrage actually to take place, it would help close the price gap, effectively stabilizing the relationship between JIF, authorship price, and bonus cash rewards. I believe this would finally provide a tangible value (though still a local and temporary "spot price") for that most elusive of goods — impact. It would also suggest that money is neither a proxy for impact nor a convenient means for rewarding it, but may, instead, have become the only unit of measurement that can truly capture it. There is nothing natural, however, to the fit between impact and money. Rather, the development of an academic economy based on impact has quite simply and effectively turned impact into its monetary currency.

Because we cannot clearly define and measure impact in and of itself, we pick quantifiable proxies like citations or derivative indexes like the JIF and then compare them with those of other publications or journals. (Metrics of impact gain whatever meaning they can gain only when structured into rankings. [20]) But this can only produce relative value. The best we can do is to say that, whatever impact may be, *Nature* has more of it than, say, the *Memorie dell'Accademia dei Lincei*. The hypothetical example of the arbitrage involving the journal impact factor suggests, however, that while a conceptually robust definition of impact and its measurement is unattainable, we may still be able to measure the *impact of impact*, or the *price of impact*. We could keep doing what we have been doing all along: use proxies for impact — money in this case — based on the recognition that JIF-indexed cash bonuses and authorship sales already exist. That proxy has already been adopted by the actors.

This does not mean that the proponents of metrics were always right but just could not prove it. It certainly does not mean that impact has always been a real thing, and that we can now measure it — finally and precisely. Instead, I would say that the introduction of basic metrics of evaluation like citation counts has spawned the articulation of various indexes meant to simplify

and standardize that job, and that some such indexes — like the JIF — have become "products" in and of themselves. Impact remains as undefined as ever, but products stemming from the processing and packaging of proxies of impact have developed an active market of their own, effectively rendering moot the problem of the definition of impact. The question of what impact is has simply fallen off the table, displaced by products like the JIF, themselves constituted through the blackboxing of that unanswerable question.

Forging relations, forging time

Manipulations of metrics take many forms, with new ones likely to emerge, but they all share some key traits because of the nature of the object they target: impact. Metric indicators work comparatively, showing a publication's impact relative to that of other publications. As a result, their manipulation does not aim at hitting a specific target — like, say, a hypothetical global warming denier massaging data in order to claim that average temperatures have not increased but actually declined by 1.5 degrees Celsius since the Industrial Revolution. With metrics manipulation there is no way to know when the books have been cooked enough. That means that one can only strive to produce *more* proxies of impact: more citations, publications in higher impact journals, higher visibility, etc. It is an inherently inflationary type of misconduct, which is yet another significant problem in and of itself.

In addition to being relative, impact is also relational in the basic sense that your article needs to be cited by another article. (Even if you want to self-cite — which is easily done but also easily detected — you need to write another paper to do that.) Traditional fraud could take place exclusively within the boundaries of one publication, but because impact is a connection between two or more publications, metrics-oriented misconduct aims at faking more of those relations (as in the case of coerced citations or citation rings) or falsely eliciting them (by manipulating, for instance, the publication process in order to place an article in a higher

impact factor journal, or by improving its visibility with fake high-power co-authors). Unless you hack into Clarivate Analytics's Science Citation Indexes to improve your numbers — a very direct approach which does not seem to have succeeded yet — you need to mobilize other publications to deliver the desired citations to your article. Metric misconduct is, therefore, ultimately about manipulating a publication's kinship network, adding fake links to it like a falsified family tree.

This crucially affects the temporal framework for the production of value. Traditional scientific fraud is an event that happens over a finite period of time, ending before the publication of the fraudulent article, but postproduction manipulations can be an ongoing process. If you are an unethical peer reviewer (or one who knows one), you could keep extracting citations to articles you published 20 years ago. Who knows, that could render your profile impactful enough to be noticed by important prize committees. Truth might be fixed in time, but impact is a never-ending game that can be played at any time. "Impact" just keeps on giving, in other words.

Time plays only an incidental role in traditional fraud, which is a one-off manipulation aimed at generating credit from the fraudulent publication right there and then. By contrast, time is central to the metrics economy and postproduction misconduct because their object — impact — is literally the product of accumulation, which can be manipulated in different ways over time, as we have seen. But the same cumulative nature of impact that provides such a wide window of opportunity for its manipulation creates another more direct problem for science by slowing down the evaluation of publications and the careers that depend on it. Impact takes a long time to build, creating a bottleneck in a publication economy that has become all about speed. Electronic publishing has significantly cut down on the length of review, production, and dissemination processes. It has facilitated collaboration, largely erased the effect of geographical

distance, and greatly increased the speed of literature searches. It is difficult to imagine the current output of science — about three million articles a year — without electronic publishing.

At the same time, since the introduction of metrics, the valuation process has been hitched to impact, which takes years to build compared to the much quicker traditional process of qualitative evaluation based on people reading their colleagues' publications and judging them. An airplane riding on a steam train, so to speak. This should not be surprising. Introduced by early Scientometrics, impact has become so pivotal precisely because, unlike the qualitative judgment of peers, it is quantifiable and can be presented as objective — it tracks the empirical traces of responses to that publication, not opinions about it. But while opinions can be generated quickly, it takes significant time for citations to be generated, recorded, and accumulated.

There is no way around it. One can argue about using different proxies to measure impact, but nobody has been able to come up with metrics without impact or impact-like concepts that, in one way or another, try to quantify the effects of a publication's reception. Impact is what makes science metrics possible, which in turn allows — so their proponents claim — a fairer and more efficient management of the very large, complex, and globally distributed behemoth that is science today. At the same time, the inherent slowness of impact creates bottlenecks in the system it has made possible — bottlenecks that are constitutive, not accidental and removable.

The emergence of Altmetrics, the success of the JIF, and the purchasing of fake impact factors are all responses — some legal, others not — to this predicament. Altmetrics was introduced around 2009 to fill the gap created by the absence of early evidence for impact prior to citations. [21] It does so by tracking other empirical traces like tweets, mentions on Reddit, blogs, and so on. Its basic logic is thus the same as that of academic metrics but using different proxies. However, this ability to operate on evidence other than citations also limits the

information it can provide. Altmetrics does provide quantitative indicators, but those indicators could in fact be measuring gossip. This is not meant disrespectfully but to suggest that, when impact metrics try to capture the early life of a publication by tracing something other than citations, then they turn into something else altogether. Altmetrics has tried to fill in a blind spot plaguing academic metrics, but it is a gap that cannot be filled without exiting the logic of, and justification for, academic metrics. (While Altmetrics may provide valuable marketing information for publishers, it is not the kind of evidence of impact that sways, say, Nobel Prize committees.)

The widespread off-label use of the JIF is motivated by similar concerns. The sleight of hand that has made the JIF as important as it is — thanks to the de facto attribution of the journal impact factor to the articles it publishes — is literally a way to shift the time axis of impact, turning the past impact of journals into the future impact of articles (that do not have impact yet), thus shaving years off the accumulation of value. While most people would agree that purchasing a fake impact factor to jump-start your new journal is or should be illegal, nobody I know would say that the off-label use of the JIF amounts to misconduct. The JIF may well provide a patch to a constitutive problem in the metrics system, but the way it is being used subverts the very logic of that index while also flipping the meaning of impact upside down. Unlike Altmetrics, the JIF has managed to reduce the drawbacks of impact-based metrics from within, but only by gaming the foundations of the whole system, thus effectively hacking time.

Content is not what it used to be

Citations keep attaching to an article over time, like little tentacles that latch on to its surface but do not reach into its interior. This happens because, in the metrics economy, the content of the article has become less important than its metadata. Evaluation does not start with reading but with the identification of an article as a bibliographical object by means of the name of the

author(s), the journal, the title of the article, the publication date, and other metadata. (Think of the metadata as the GPS coordinates of the publication.) Once the object has been identified and located, algorithms then trace the citations it has received over a certain period of time, the impact factor of the journal in which it appeared, its relative impact ranking compared to other articles, and so on. (It is telling that metadata has grown more extensive and granular over time to facilitate this accounting work. [22])

The article needs to have content, and yet that content does not function as the focus of evaluation. People still read for research and educational purposes, but reading is no longer a necessary component of institutional forms of evaluation because metrics are independent of the epistemic dimensions of that specific publication — its claims — but rely, instead, on metadata and similar markers that can be picked out and processed by nonhumans. The article has become more like a vector and recipient of citations than a medium for the communication of content. The content needs to exist, but as a hook to attach the metadata to. The evaluation hinges only on the "envelope" of the article, but there needs to be content to justify the envelope rather than the other way around. The metadata has to be the metadata of something. To use a different figure, think of the content as providing a target. The only thing that matters are how many arrows hit the target, but there can be no hits without a target.

Recycling fraud

All this does not mean that traditional misconduct — fabrication, falsification, and plagiarism — has been rendered obsolete. On the contrary, it, too, is the site of innovation. In particular, Photoshop has turned image manipulation into a labor-saving, computer-aided form of scientific fraud, making it possible to crop, delete, splice, compress, stretch, tilt, flip, and recombine bands, and modify their contrast level to produce fraudulent but professional-

looking images. (Figure 1) Most cases concern blots like these, but any kind of photographic evidence may be involved.

Mike Rossner and Kenneth M. Yamada, "What's in a Picture? The Temptation of Image Manipulation" The NIH Catalyst 12: 8–11 (May–June 2004)
http://www.nih.gov/catalyst/2004/04.05.01/page4.html

These images may then be used as evidence not just in one article but in several, each one recycling the same manipulated blot to falsely reference different experimental conditions and samples. More than 60 percent of the 2018 misconduct findings by the US Office on Research Integrity involved image manipulation.

Photoshop has increased fraudulent output and, more importantly, changed the very form of traditional, epistemic fraud. Older fraud involved unique, local, hand-manipulated evidence crafted to support specific claims. The 1912 Piltdown Man case, for instance, involved planting

physical evidence: medieval human craniums were mixed in with modern orangutan jawbones, physically manipulated and then placed in a pit to create evidence of the "missing link" between humans and apes. Similarly, in 1926, Paul Kammerer was accused of physically tattooing "nuptial pads" on midwife toads in his laboratory in Vienna to support Lamarckian against Darwinian evolution, and in 1974 William Summerlin admitted to using a felt-tip pen to color a skin graft on an albino mouse at Sloan-Kettering in New York to claim that he had been able to suppress the immunological rejection of transplanted tissues. Now, however, fraud has moved from the field or laboratory to the computer. Its mode of production remains artisanal, but digital tools allow for quasi-industrial productivity.

It is important to realize that digital tools make the *reproduction* of fraud — not just its production — virtually effortless. An example is the recycling of identical or slightly modified fraudulent blots to support different claims in different articles. In one instance, the same blot was identified in 15 images across 10 different articles. [23] Categorically new, *recursive fraud* of this kind is the serial digital reproduction and dissemination of more fraud based on tweaks or copies of previous fraudulent evidence. An art forger who paints a "lost Rembrandt" will try to pass it off as one authored by the Dutch master, which it is not. Still, that forgery remains a unique painting original to the forger. Similarly, Summerlin's painted mice may not have been works of art, but they were still handmade originals, not copies. They were *fraudulent and yet authentic*. By contrast, today's serially photoshopped images are *fraudulent copies of fraudulent copies*. They are re-manipulations of previous manipulations or, in a few cases, non-fraudulent images that are copied and repurposed for fraudulent ends, with fake captions.

The technologies that make it possible to disseminate multiple fraudulent copies instead of just one singular authentic fraud are precisely what make the new digital fraud so productive, which brings us back to metrics. If it were not for an obsession with productivity, why would we find

the same fraudulent blot reproduced, more or less identically, 15 times in 10 publications? It seems that these scientists are in such a rush that they take shortcuts *even within fraud-making*, repurposing fraudulent evidence because that is much faster than faking it from scratch, as their predecessors did in the pre-digital age.

Digitally manipulated images look professional and thus prima facie credible. But that thin veneer of credibility lasts only until readers or editors probe a little harder. People with good eyes and some training can spot duplication and manipulations, and there is now software for detecting more sophisticated manipulations. [24] Unlike traditional fraud where forensics could be lengthy and complex, digital fraud is likely to crumble like a house of cards shortly after a red flag is raised.

What we tend to see today are not ambitious fraudulent publications that try to escape detection thanks to their sophistication. The recursive fraud discussed here is more akin to making many sloppy knockoffs of Levi's blue jeans than one masterful forgery of Leonardo. In the age of metrics, quantity trumps quality even in fraud. Rather than investing time and effort in developing hard-to-detect fabrications of important claims, fraudsters use standard digital tools to produce many articles that, while not likely to withstand much critical scrutiny, may survive in lower-tier journals where editors are less exacting and reading is optional.

¤

The two parts of this essay — one about postproduction misconduct and the other about innovations in traditional fraud — describe two regimes of manipulation that seem to have virtually nothing in common. The former is indifferent to the content of a publication and manipulates only its metrics. The latter is about manipulating a publication's visual evidence while leaving its metrics untouched. This neat complementarity becomes blurred in practice,

when the manipulation of metrics can coexist with the manipulation of content. There are also two overlaps of a more conceptual kind.

The first: Both forms of misconduct are framed by metrics, even when metrics are not explicitly manipulated. Photoshop-fueled, high-throughput fraud is aimed at meeting and exceeding productivity benchmarks. In that case, metrics are manipulated by manipulating the publications themselves and their quantity rather than other external indicators.

The second: Both approaches disregard the traditional assumption that content is for reading. Photoshopped, recursive fraud produces false images (which may then be republished several times) that can easily be exposed as fraud. Postproduction fraud, instead, simply disregards the publication's content and only manipulates its envelope. In both cases, readers and reading seem to be irrelevant. The low detection threshold of Photoshop fraud suggests that those who practice it probably see readers as dangerous, and hope to gain as much credit as possible for their high productivity before they get caught. Credit comes from the number of publications, and so readers are a liability. Postproduction misconduct goes a step further, casting readers as effectively irrelevant. Credit comes from manipulating metrics, and the readers that matter are the algorithms that gather the traces of one's fake impact and turn it into scores one can use.

¤

Click here to listen to a podcast conversation between Mario Biagioli and John Markoff on fraud and gaming in scholarly publishing.

¤

Mario Biagioli is a professor in UCLA's School of Law and its Department of Communication. He is the author most recently of two books, From Russia with Code: Programming Migrations in Post-Soviet Times *(co-edited with Vincent Lepinay, Duke University Press, 2019) and* Gaming the Metrics: Misconduct and Manipulation in Academic Research *(co-edited with Alexandra Lippman, MIT Press, 2020).* *This article was made possible by a fellowship at Stanford's Center for Advanced Study in the Behavioral Sciences.*

¤

[1] Cathy O'Neil, *Weapons of Math Destruction* (Crown, 2016), pp. 50–67.

[2] Max Kutner, "How to Game the College Rankings," *Boston Magazine*, August 2014.

[3] Bill Readings, *The University in Ruins* (Harvard University Press, 1996), pp. 23-43.

[4] Mario Biagioli, "Quality to Impact, Text to Metadata," *KNOW* 2 (2018).

[5] Marilyn Strathern (ed), *Audit Cultures* (Routledge, 2000); Michael Power, *The Audit Society* (Oxford University Press, 1997).

[6] Mario Biagioli and Alexandra Lippman (eds), *Gaming the Metrics: Misconduct and Manipulation in Academic Research* (MIT Press, 2020), pp. 1–3.

[7] Adam Marcus, "Journal retracted at least 17 papers for self-citation, 14 with same first author," *Retraction Watch*, January 6, 2020.

[8] Alberto Baccini et al., "Citation gaming induced by bibliometric evaluation," *PLoS*, September 11, 2019.

[9] Dalmeet Singh Chawla, "Elsevier Probes Dodgy Citations," *Nature*, Vol. 573, September 12, 2019, p. 174

[10] Adam Marcus, "Author appeared to use phony Caltech co-authors, up to 8 retractions," *Retraction Watch*, April 7, 2016.

[11] Junhui Han and Zhengfeng Li, "How Metrics-Based Academic Evaluation Could Systematically Induce Academic Misconduct: A Case Study," *East Asian Science, Technology and Society* (2018) 12 (2): 165–180.

[12] "Bonuses Encourage Lecturers to Do Scientific Research, but Problems Remain," *Viet Nam News,* January 6, 2020: "[S]ome lecturers were naming other universities in their work to get higher bonus pay-outs."

[13] https://liorpachter.wordpress.com/2014/10/31/to-some-a-citation-is-worth-3-per-year.

[14] W. Brian Arthur, *Complexity and the Economy* (Oxford University Press, 2014), pp. 103–118.

[15] John Tregoning, "How Will You Judge Me if not by Impact Factor?" *Nature*, Vol. 558, June 21, 2018, p. 345.

[16] Biagioli and Lippman, *Gaming the Metrics*, p.13.

[17] Donald MacKenzie, "Is Economics Performative? Option Theory and the Construction of Derivatives Markets," *Journal of the History of Economic Thought* 28 (2006): 29–55.

[18] Alison Abritis et al., "Cash Incentives for Papers go Global," *Science*, August 11, 2017: Vol. 357, Issue 6351, pp. 541.

[19] Having grown more confident with the impact of its publications, China has just banned these practices, Smriti Mallapaty, "China Bans Cash Rewards for Publishing Papers," *Nature*, Vol 579, March 5, 2020, p. 18.

[20] Biagioli, "Quality to Impact," pp. 13–15.

[21] Jennifer Lin, "Altmetrics Gaming: Beast Within or Without?," in Biagioli and Lippman (eds), *Gaming the Metrics*, pp. 213–227.

[22] Yves Gingras, "The Transformation of the Scientific Paper" in Biagioli and Lippman (eds), *Gaming the Metrics*, p. 47.

[23] Leonid Schneider, "Mario Saad and the Return of the Wandering Western Blot," *For Better Science*, June 21, 2016.

[24] Elisabeth M. Bik, Arturo Casadevall, Ferric C. Fang, "The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications," *MBio*, May/June 2016 Volume 7, Issue 3, p. 5.

Mario Biagioli

Mario Biagioli is a professor in UCLA's School of Law and its Department of Communication. He is the author most recently of two books, *From Russia with Code: Programming Migrations in Post-Soviet Times* (Duke University Press, 2019) and *Gaming the Metrics: Misconduct and Manipulation in Academic Research* (MIT Press, 2020).