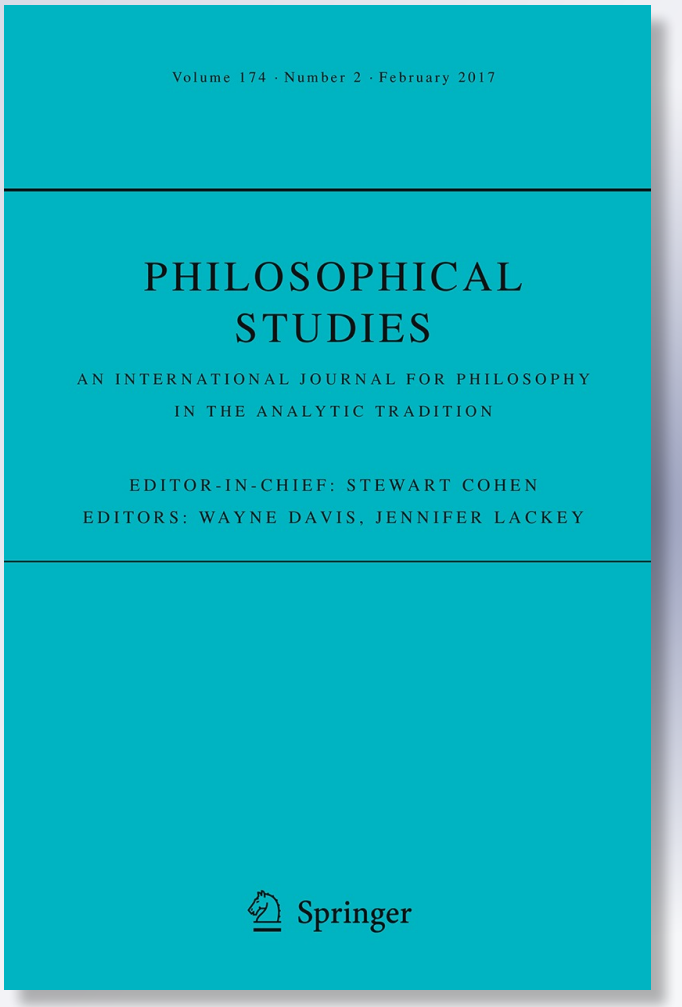
On fraud

Liam Kofi Bright

Philosophical Studies
An International Journal for Philosophy in the Analytic Tradition

ISSN 0031-8116
Volume 174
Number 2

Philos Stud (2017) 174:291-310
DOI 10.1007/s11098-016-0682-7



Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Author's personal copy

Philos Stud (2017) 174:291–310
DOI 10.1007/s11098-016-0682-7

On fraud

Liam Kofi Bright[1]

Published online: 29 April 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Preferably scientific investigations would promote true rather than false beliefs. The phenomenon of fraud represents a standing challenge to this veritistic ideal. When scientists publish fraudulent results they knowingly enter falsehoods into the information stream of science. Recognition of this challenge has prompted calls for scientists to more consciously adopt the veritistic ideal in their own work. In this paper I argue against such promotion of the veritistic ideal. It turns out that a sincere desire on the part of scientists to see the truth propagated may well promote more fraud rather than less.

Keywords Philosophy of science · Social epistemology · Decision theory · Fraud · Social structure of science · Veritism

1 Introduction

Preferably scientific investigations would promote true rather than false beliefs. The phenomenon of fraud represents a standing challenge to this veritistic ideal. When scientists publish fraudulent results they knowingly enter falsehoods into the information stream of science. Recognition of this challenge has prompted calls for scientists to more consciously adopt the veritistic ideal in their own work. In this paper I argue against such promotion of the veritistic ideal: a sincere desire on the part of scientists to see the truth propagated may well promote more fraud rather than less.

✉ Liam Kofi Bright
lbright@andrew.cmu.edu

[1] Department of Philosophy, Baker Hall 161, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Author's personal copy

To understand the background of such calls for scientists to be more consciously veritistic, it is necessary to understand the role of the credit motive in incentivising scientific work. A significant proportion of the activities of scientists can be explained by appeal to the fact that they seek to maximise their credit (Cole and Cole 1967; Merton 1968; Dasgupta and David 1994; Latour and Woolgar 1986, ch. 5). It is well established by this point that in order to understand a group's ability to produce and disseminate knowledge we must understand facts about how that group is socially organised (Zollman 2010; Muldoon 2013; Kummerfeld and Zollman, 2015). As such, social epistemologists have studied the credit economy, and have largely found that such selfish credit seeking behaviour by scientists inadvertently promotes socially beneficial epistemic outcomes (Kitcher 1990; Dasgupta and David 1994; Strevens 2003, 2006; Bruner 2013; Muldoon 2013, §2). This research programme has thus been described as an 'invisible hand' tradition, since the basic idea is that while decision making at the individual level may not be guided by the desire to ascertain the truth, this tends to be conducive to truth finding success at the social level (Solomon 2001, 55).

However, running contrary to this happy social epistemic consensus, the phenomena of fraud has been claimed by many sociologists to be a negative consequence of the operation of the credit economy (Merton 1973, 309–312; Ben-Yehuda 1986, 5; Zuckerman 1988; Sovacol 2008, 275–277; Casadevall and Fang 2012, 892). For instance, drawing a moral from examining a case study of scientific fraud, Broad and Wade conclude: "Many scientists, no doubt, still keep truth as their goal. For many, however, a more immediate objective often intrudes into vision, that of establishing credit" (Broad and Wade 1983, 52–53). It is this more immediate objective that they believe leads to fraudulent publication. Likewise Fanelli et al. (2015) claim the motivation for committing fraud is "to gain an unfair advantage in the race for priority and success". Given that fraud involves the deliberate introduction of error into the epistemic community, this theory would require the invisible hand tradition to modify its epistemic estimation of the credit economy.

This theory about the cause of fraud has lent itself to policy proposals to modify or eliminate the credit economy, replacing it with more conscious adoption of the veritistic ideal. I refer to works that advocate such policies as the 'motive modification' tradition of anti-fraud research. The following quote succinctly summaries the ambitions of the motive modification tradition: "[t]he persistence of false findings can be meliorated with strategies that make the fundamental but abstract accuracy motive – getting it right—competitive with the more tangible and concrete incentive—getting it published" (Nosek et al. 2012, 615). Some have even gone further, with Du Bois (1898) arguing that the only acceptable motivation for scientists is the pure pursuit of the truth, which would have to be unsullied by concern for credit. However one brings about motive modification, optimistically, one might even hope that a suitably designed incentive structure could retain the benefits of the credit economy mentioned above while also using the tempering effects of a truth motive to simultaneously reduce the incident of fraud.

The overarching lesson of this paper, however, is that motive modification is not enough. This is because the theory that it is based upon is in need of update or

Author's personal copy

revision. The identification of the credit motive as the source of scientific fraud represented an overly simplistic understanding of the causes of fraud. In Sect. 2 (and in formal detail in the "Appendix") I describe a model of academic publication which allows me to investigate counterfactually what behaviors to expect given different motivations among scientists. In Sect. 3 I report some results (proven in the "Appendix") that undermine the motive modification tradition of anti-fraud research. Simply making truth compete on equal footing with credit only changes the situations in which scientists are incentivised to commit fraud, but does nothing to suggest that scientists will be incentivised to commit fraud less often. Further, by considering the results of scholars assuming there is some cost to committing fraud, I show that pure credit seekers will in some circumstances be incentivised to report just those results they obtained while those who are concerned with the truth are incentivised to do otherwise. In Sects. 4, 5 and 6 I discuss these results in relation to both the motive modification tradition of anti fraud and also work elsewhere in social epistemology of science. I argue that what explains the failure of the motive-modification tradition in anti-fraud research is that the desire to promote true beliefs among one's fellows can incentivise a scholar to lie when the scholar distrusts the results of the experimental work they themselves have carried out.

2 Publication market models

To investigate the efficacy of motive modification anti-fraud policy, I investigate the behaviour of agents within a class of models I call *publication market models*. These represent an arbitrarily large field of scholars trying to discern the truth of the matter regarding some central question of interest. The formal details of publication market models, along with relevant proofs, can be found in the "Appendix". I give informal description of publication market models here.

Within any publication market model, there is a set of scholars, and a question represented by a set of possible answers. Each scholar does some investigation on the question, and comes to favour one of the answers. Following (Hintikka 2007, ch. 2), inquiry is modeled as scholars asking a question of Nature and receiving some answer. In the model I represent Nature as answering questions by sending each scholar an independent and identically distributed signal of one of the answers from within the question set. The signal received by a scholar raises their credence in whatever answer it suggested. There is one answer that is epistemically favoured—the "true" answer—regarding the question. In the model truth is represented by Nature selecting an answer in advance of distributing signals.

To represent publication, after Nature has distributed signals among the scholars, each scholar publicly announces some answer to the question, that need not be the one Nature signaled to them. Note that the signals are private to each scientist, what is public and presumed to be shared among the scientists are answers to the question. The model hence does not suppose that each scientist has applied the same method to the same question, and that should all come to the same answer we may have indefinitely many scientists publishing essentially identical papers. Rather, this is intended to model of the following sort of situation. There is some question of

Author's personal copy

common interest to many in a field, e.g. what caused the extinction of the dinosaurs? Can a material with certain properties be cheaply synthesised? Does raising the minimum wage increase unemployment? The scientists in the field share a sense of what the possible answers to this question are, and are able to identify when two different investigations have come to the same answer on this question. For instance, they are able to recognise data sets drawn from different times and regions which both show no rise in unemployment following some law raising the minimum wage as both supporting a 'No' answer to the question: does raising the minimum wage increase unemployment? I assume it is answers to such questions, rather than raw presentations of data, which are the objects of scientific interest, and the potential sources of scientific credit.

That question about unemployment represents an easy case because it leads to a simple 'Yes/No' dichotomy in its possible answers. Often it will not be so easy to state precisely what answers may be expected to a question in advance of investigation. This model of scientific investigation hence represents an idealisation of aspects of scientific practice highlighted by Kuhn (1996, ch.4). Kuhn argues that, when engaged in normal science, scientists who share a paradigm will agree on 'the nature of acceptable solutions' to the puzzles they try to solve (Kuhn 1996, 38). I idealise this picture by modelling such agreement to go so far as to yield identical sets of possible answers to the question being investigated. This idealisation does not play too great a role in what follows, however, as the results I establish and rely upon concern the reasoning of an individual scientist trying to anticipate the results of their fellows' investigations. Nothing would be lost if it turns out such a scientist did not accurately portray the range of possible answers or did so differently from a peer, as their personal estimation of the range of possibilities is what matters for their own publication decisions.

Once agents have published there is an election among the answers that have been announced, with the winner of the election representing the answer(s) to the question that the community has come to settle upon. The election over announced propositions represents uptake within target communities that a scholar may care about when deciding what to publish. Note that I make minimal assumptions about how the winner of the election is decided. I assume only that being more popular cannot count against an answer; that if a given answer would have been consensus had one less person opted for it, then all else being equal it will still be victorious if it receives that extra vote. To illustrate the sort of uptake I have in mind, I mention here uptake within two sort of target communities the scholar may be concerned with. First, the election over announced propositions may represent a sort of bounded rationality within an academic scholarly community. Academics may not be able to keep track of the precise degree to which various answers are supported within a scholarly literature, but they can at least review the literature and decide what response to the question is best supported by their results of the community's collective endeavour. Second, the election over announced propostions may represent uptake among a non-academic audience (e.g. policy makers, or representatives of industry) who are not competent to review a literature in depth but are able to gain a general sense for what propositions are best supported therein.

Author's personal copy

Each scholar has a *utility type*; differences between utility types allow for the investigation of the effects upon fraudulent publication of different incentive structures. I explore three utility types in this paper. The first is a *pure credit seeker*, who cares only that the community comes to believe whatever answer to the question the credit seeker themselves has announced, since the credit seeker's scholarly reputation is associated with the fate of this position. In the actual credit economy of academia in order for a scholar to gain significant credit from a publication a scholar does not just need the community to settle upon their favoured proposition but also need to be seen to be first to arrive at that proposition (Strevens 2003). However, note that in that scenario it is still a necessary condition for any credit seeker to gain their reward that the community adopts their preferred belief. Further, it is not always clear who will be awarded priority for a given claim, so a credit seeker may justifiably give positive subjective probability to the event that it will be them, as long as they defend the proposition that the community ends up agreeing on. The second utility type is that of the *pure truth seeker* who cares only that the community consensus—the victor of the election—is whatever proposition Nature favoured, without concern as to whether they are on the right side of that consensus. Finally, there is the *mixed credit/truth seeker*, who is concerned both that the community come to believe the proposition Nature has favoured when dispensing evidence, but also that they be seen to be on the correct side of this consensus in their contribution to the literature.

Before moving on, it is worth noting the manner in which I characterised the utility types in this model. All utility types have a binary preference structure. Their rewards are either all or nothing; either the community comes to accept just what they would prefer and they win, or they get no payout at all. One might consider an alternate way of representing truth tempering the operation of the credit economy by means of scholars with more nuanced preference structure. In particular, to capture the idea of tempering the credit economy by considering scholars who do want to win credit, but who if they are not to get credit would at least prefer the community to settle upon the truth. However, in lemma 2 of the "Appendix" I show that this alternate representation of nuancing the operation of the credit economy would not work, since such a scholar would behave exactly like a pure credit seeker.

Scholars also have a *credence function*. This is a probability distribution over what I dub each scholar's *relativised state space*. This is a construction each scholar generates for themselves representing the possible states of the literature they think they could face at point of publication. Formal details are given in the "Appendix". For present purposes it suffices to note that the relativised state space for a given scholar represents all the ways they think their peers could publish, under the supposition that each of the possible answers in fact obtains. An example of such a relativised state space is given below.

Scholars maximise expected utility given their credences over the relativised state space and given their utility type. I say a scholar is *incentivised to fraud* just in case it would not be an expected utility maximising response to the literature to offer up the answer that Nature signaled to them. The quoted remarks in §1 from Nosek et al. suggest that their analysis of the situation regarding fraudulent publication is that presently all too many scholars are incentivised, consciously or

Author's personal copy

unconsciously, to commit fraud because they are pure credit seekers, and that anti-fraud reform can be achieved by converting these agents into mixed credit/truth seekers.

A more far reaching reform, defended by W.E.B. Du Bois, would be insisting that scholars ought to be motivated by only the desire to attain and disseminate the truth. In particular, Du Bois claimed that "[s]tudents must be careful to insist that science as such... has but one simple aim: the discovery of truth" (Du Bois 1898, 16) This was to be a public rather than private acquisition of truth, a communalist rather than individualist pure truth seeking, since Du Bois also insisted that science's "results lie open to the use of all men" before reiterating that "the aim of science as such is simple truth" (Du Bois 1898, 16). Hence scientific investigation should be, according to (as I shall call it) DuBoisian alethic puritanism, motivated only by a desire that truth should be believed by as many as possible.

The results I obtain will show that without special assumptions being made about how scholars form their credences over the relativised state space there is no reason to believe that motive modification would disincentivise fraud. To illustrate some of what is driving results in this model, I discuss two behavioural assumptions one could make about scholars, focussing on how the behavioural assumptions affect the prospects of motive modification anti fraud research.

The first assumption I call 'Cost of Fraud', since it represents scholars acting as if there is some small risk associated with committing fraud. Under this assumption scholars take there to be some small $\epsilon_1 > 0$ cost to publishing fraudulent results. This may obtain, for instance, if scholars think that there is a very low chance that they will get caught for committing fraud, but expect to be punished if they do. I consider the effects of this behavioral posit in that class of cases where scholars think that, first, how they announce will be decisive in what the community comes to believe, and, second, that the community could be made to settle upon the answer the scholar takes Nature to have favoured. The justification for singling out this class of cases is given in the next section. In the "Appendix" I show that within this class of cases Cost of Fraud more often makes credit seekers incentivised to honesty than it does for truth seeking scholars.

The second assumption I call 'Self Confidence': a scholar is self confident if they assign some small $\epsilon_2 \geq 0$ credence to any element of their relativised state space where Nature did not favour whatever signal Nature sent them. This models a scholar who is very sure of their own experimental work; they believe whatever their experiments reveal, regardless of what they take other scientists to have learned. I show there that if this is how scholars assign their credences then motive modification may serve to reduce fraud. Future work in publication market models could profitably focus on exploring the behaviour of scholars under different classes of assumptions about how they assign credences in light of the signal Nature sends them.

To illustrate the model at work, consider the following arrangement of the model. The set of scholars has three members {1, 2, 3}. They are investigating some proposition, which has two potential answers $\{a_K; a_J\}$. I construct below scholar 3's relativised state space, and tabulate the payouts 3 would receive for given announcements (with possible announcements listed on the far left hand side) were

Author's personal copy

they a pure credit seeker, mixed credit/truth seeker, or pure truth seeker respectively. The top row of this table represents Nature's favoured answer. The second row of the table represents various combinations of publications scholar 3's peers may offer to the literature. Note that each announcement is superscripted with the scholar who made it and subscripted by the answer it supports. Assume the election is majoritarian. The other rows represent the payouts scholar 3 would receive under various suppositions about their utility type and which answer they submit to the literature.

To illustrate—the second from left column represents the scenario where Nature favours a_K, and both scholar 1 and 2 publish defences of a_K. Whereas the fifth from left column represents the case where Nature favours a_J, while scholars 1 and 2 none the less both publish in defence of a_K. Whether scholar 3 would be incentivised to fraud in such a scenario would depend on how their credence function distributed probability among these possibilities, what Nature had signaled to them, and their utility type. So suppose, for instance, that they believed it overwhelmingly likely (for concreteness sake: they assign credence 1) that they are in the rightmost cell of the table. In that case if Nature signaled to scholar 3 a_K then they could be incentivised to fraud if they were a pure or mixed credit/truth seeker, but not if they were a pure truth seeker.

3 Motive modification results

I report four theorems relevant to assessing the motive modification tradition of anti-fraud research. Formal proofs are given in the "Appendix". The first result I dub the "Pessimistic theorem", since it undermines the meliorist reform programme advocated by Nosek et al. (2012), and perhaps also Broad and Wade (1983). According to the pessimistic theorem, the class of elements of a scholar's relativised state space which, if given any positive credence, will constitute some temptation to commit fraud, has the following features. The only circumstances under which converting a pure credit seeker into a mixed credit/truth seeker could reduce the scholar's incentive to fraud are within those elements of the relativised state space when the scholar does not think they can actually sway what answer the scientific community comes to accept. What is worse, converting a scholar into a mixed credit/truth seeker from a credit seeker can actually introduce incentives to lie where previously there were none. Driving that latter result are occasions where a scholar does not think their own results to be representative of the truth but does think themselves to be influential. If, for whatever reason, a scholar thinks their results are flawed, but also believes themselves capable of publishing answers to the question that will sway their fellows towards the truth, then the mixed credit/truth seeker is tempted to tell a sort of noble lie. Whereas the pure credit seeker will just as happily take the credit for the result they actually obtained, flawed or not. Note that in these scenarios it is not quite that the pure credit seeker is incentivised not to lie; they may just as happily, for all that has been said, report honest as dishonest results. Some consideration of when the scholar might be actively incentivised to tell the truth comes below. At the least, however, the proof of the pessimistic theorem shows that

Author's personal copy

sometimes the pure credit seeker has no incentive to lie where the mixed credit/truth seeker does. More generally, it shows that converting scholars from credit seekers to truth seekers could increase rather than decrease their incentive to fraud, or have no effect; absent further information, motive modification cannot by itself be considered an anti-fraud policy.

The second theorem I dub "Du Bois' Conjecture", because it is a partial vindication of his alethic puritanism. According to Du Bois' Conjecture, changing scholars' motivations so that they are pure truth seekers rather than mixed credit/truth seekers strictly decreases the class of scenarios in which a scholar is incentivised to lie. If two scholars are otherwise identical except one is a pure truth seeker and the other a mixed credit/truth seeker—then the set of elements of relativised state space wherein the latter would best respond by publishing a fraudulent answer is a proper superset of the set of elements of relativised state space wherein the former would best respond by publishing a fraudulent answer. This is a partial vindication of Du Bois' point since it supports his claim that giving science a 'double aim' is worse than keeping one's motivations purely truth focussed (Du Bois 1898, 16). However, it is only a partial vindication because, first, this theorem says nothing about the difficulties involved in implementing Du Bois' suggestion, and, second, (as can be seen in Table 1) there are still situations under which a pure truth seeker might be incentivised to commit fraud.

The third theorem I report I dub "Active Honesty", since it is a consideration of cases where scholars will take honest reporting to be their unique expected utility maximising option. One apparent result of the Pessimistic theorem was that in situations where a scholar does think they can sway the scientific community motive modification research can make no difference. I note that under the Cost of Fraud assumption this is not strictly speaking true. Suppose scholars obey Cost of Fraud and consider those elements of the state space where, first, scholars believe that how they vote will make a difference to what the scholarly community comes to accept, second, the scholar's preferred option is among the beliefs the scholar supposes the community could come to accept, and, third, the scholar believes they could induce the community to endorse Nature's favoured option. Theorem three shows that in such cases pure credit seekers' best response will always be to report honestly, whereas the mixed or pure truth seeking scholar can be incentivised to

Table 1 Scholar 3's relativised state space

Nature's choice:	a_K	a_K	a_K	a_K	a_J	a_J	a_J	a_J
The literature:	$a_K^1; a_K^2$	$a_K^1; a_J^2$	$a_J^1; a_K^2$	$a_J^1; a_J^2$	$a_K^1; a_K^2$	$a_K^1; a_J^2$	$a_J^1; a_K^2$	$a_J^1; a_J^2$
Pure credit: a_K^3	1	1	1	0	1	1	1	0
Pure credit: a_J^3	0	1	1	1	0	1	1	1
Credit and truth: a_K^3	1	1	1	0	0	0	0	0
Credit and truth: a_J^3	0	0	0	0	0	1	1	1
Pure truth: a_K^3	1	1	1	0	0	0	0	1
Pure truth: a_J^3	1	0	0	0	0	1	1	1

Author's personal copy

commit fraud. This is because the pure credit seeker only wants to be on the winning side, and if there is a cost to fraud then they should prefer to win without paying that cost, so will bring it about that the announcement their data honestly favours is the winning option. A truth seeking scholar, on the other hand, has nothing to gain from having the community accept their favoured option if they do not think the answer their own results favour is the option Nature favoured. As such they would be willing to lie under that scenario where the credit seeker would not.

The fourth theorem I report I dub the "No Insecurity" theorem, since it represents the results of simultaneously inducing motive modification and making scientists extraordinarily confident in their own results. Under the previously mentioned assumption of Self Confidence, if Nature sends a scholar a signal favouring a given answer to the question, they assign small $\epsilon_2 \geq 0$ credence to Nature favouring any answer other than that suggested by their own signal. The scholar is maximally confident in the reliability of their own signal. No Insecurity shows that while the pure credit seeker could still be incentivised to commit fraud under such a scenario, the pure or mixed truth seeking scholar could not be incentivised to commit fraud should they also satisfy self confidence. Discussion of why this theorem holds will follow in Sect. 5.

4 Against truth

The results of the publication market model suggest a need to rethink the motive modification tradition of anti-fraud research. First, both the pure truth seeker and the mixed/credit truth seeker can be incentivised to lie even in scenarios where the pure credit seeker would not be. That is to say, the truth motive can actually introduce new reasons to commit fraud where there was none under the credit incentive. Consider, for instance, the third column from the left in Table 1, on the supposition that the agent received signal a_J from Nature. Both the pure truth seeker and the mixed credit/truth seeker would do best to announce dishonestly in that scenario, whereas the pure credit seeker could happily announce the signal they actually received. Even more dramatically, suppose that Cost of Fraud holds. Then, in an instance of the general point made by Active Honesty, in this column the pure credit seeker's unique best reply is to be honest, whereas truth seeking scholars would have a unique best reply in announcing dishonestly.

Second, where the desire for the community to reach the truth reduces the incentive to lie, it does so only in scenarios where scholars do not think themselves able to influence the beliefs of their peers. To invest in motive modification anti-fraud policy is, therefore, to bet against the vanity of scholars: it is to invest resources in a policy that could only have its desired effects on the assumption that credit seeking scholars often do not think themselves able to sway their colleagues. If one supposes that scholars' are tolerably good at estimating how likely it is that their publications will sway the community for or against some position, this latter fact introduces a further problem for the motive modification tradition of anti-fraud research. It would then be the case that the only time it is likely that the difference between credit seeking and truth seeking makes a difference are those scenarios

Author's personal copy

where a scholars' fraud is likely to do least damage to the enterprise of science considered on a social level.

Considering Active Honesty exacerbates this worry about the relationship between truth seeking and decisiveness. While motive modification anti-fraud policy may only make a positive difference when scholars do not believe themselves to be decisive, it can be actively harmful in scenarios where scholars do take themselves to be decisive. Motive modification anti-fraud policy would seem to be targeted at the most harmless sorts of fraud, and may reduce the incentive to be honest at just those points where dishonesty could be most harmful.

Combining the Pessimistic and Active Honesty theorems we thus see: truth seeking is liable to make the most epistemically powerful, those most able to sway the opinion of others, more likely to commit fraud. The scientific community is in actuality highly stratified (Cole and Cole 1973), meaning that there really are some people with an especially large influence over the opinions of others in the community. It is these high status individuals, whose errors are most likely to propagate throughout the community, that the model suggests would be more likely induced to commit fraud by the truth motive. Assuming that high status individuals recognise their own social position, the model suggests that widespread motive modification is very risky in our stratified social environment.

That said, if one must carry out motive modification anti-fraud research, Du Bois' Conjecture gives some reason to suppose that making scientists into pure truth seekers is a better idea than making them into mixed credit/truth seekers. Take, for instance, the case represented by Table 1, there one finds that mixed credit/truth seeker can actually be incentivised to dishonesty in more elements of their relativised state space than an equivalent pure credit seeker would be. Whereas this does not hold for the pure truth seeker. Du Bois' Conjecture tells us that, more generally, the class of scenarios which might induce a pure truth seeker to lie is a proper subset of the class of scenarios that might induce a mixed credit/truth seeker to lie. However, even pure truth seeking is not enough to, by itself, guarantee a reduction in the class of scenarios wherein scholars are incentivised to commit fraud. Once again, Table 1 illustrates a case wherein the pure truth seeker is incentivised to commit fraud in just as many elements of their relativised state space as the pure credit seeker.

The lesson I hope to come out of the Pessimistic theorem, Active Honesty, and Du Bois' Conjecture is that motive modification by itself is not enough. Without information about which sort of scenarios scholars consider likely one cannot know whether changing their utility functions will reduce fraud. For all we presently know it is possible that motive modification could render fraud more tempting rather than less.

5 Noble lies

Analysis of the model suggests a reason for the failure of the motive modification programme. The motive modification anti-fraud tradition has failed to pay sufficient attention to the possibility of 'noble lies'. These are cases of the desire to have the community come to believe true propositions forming its own incentive to lie, because one feels that one's own evidence would lead the community astray. The

Author's personal copy

Active Honesty theorem illustrates what a difference such a desire can make. In situations where the pure credit seeker has no reason to pay the cost associated with committing fraud, the truth seeking scholar has a countervailing incentive. The desire to see the truth propagated can make scholars willing to pay the cost associated with fraud, taking on a risk to their personal career for the sake of what they believe to be the communal good.

The No Insecurity theorem supports this analysis, and in doing so shows that there may be hope for the motive modification tradition of anti fraud research if it is also possible to eliminate the conditions which give rise to noble lies. To see this, consider why the No Insecurity theorem holds. Suppose that in the scenario depicted by Table 1, Nature sent signal a_K to scholar 3. Suppose they were self confident in the way described, and remove from Table 1 all elements of the state space Self Confidence tells us scholar 3 assigns very low credence to. Their relativised state space would then look as such:

Note that whether scholar 3 is a pure or mixed truth seeking scholar, announcement a_K—their honest announcement—weakly dominates a_J in Table 2. This is because the only scenarios where a truth-seeking scholar is not indifferent between their announcements are ones wherein the signal Nature sent them is preferred. What this shows is that self confidence renders the scholar very confident that the only way for the community to get at the truth is for it to come to agree with the results of their own experiments. As such, there are no noble lies for such scholars.

It is difficult to gather from observation of scientists' present behaviour how often noble lies are uttered. We currently operate within a credit economy where personal and professional rewards come with publishing one's results. As such it is difficult to disentangle the desire for personal reward from the desire to promote one's favoured beliefs, even in cases where it is clear the fraudulent scholar in question believes wholeheartedly the claims they fraudulently support.

There is, however, historical evidence that truth seeking scholars have been incentivised to produce what we would now think of as noble lies. In particular, this is illustrated by the behaviour of those who adhered to the 'truth-to-nature' epistemic ideal as described in (Daston and Galison 2010, ch.2). Daston and Gallison describe the behaviour of scientific atlas compilers, engaged in the project of trying to create images of objects of interest for the use of naturalists working in

Table 2 Scholar 3's restricted relativised state space

Nature's choice:	a_K	a_K	a_K	a_K
The literature:	$a_K^1; a_K^2$	$a_K^1; a_J^2$	$a_J^1; a_K^2$	$a_J^1; a_J^2$
Pure credit: a_K^3	1	1	1	0
Pure credit: a_J^3	0	1	1	1
Credit and truth: a_K^3	1	1	1	0
Credit and truth: a_J^3	0	0	0	0
Pure truth: a_K^3	1	1	1	0
Pure truth: a_J^3	1	0	0	0

Author's personal copy

fields like botany or crystallography. Images of the objects of interest would be produced based upon observed samples; however, atlas creators would not reproduce precisely what was observed, for the concern was that 'nature is full of diversity, but science cannot be' (Daston and Galison 2010, 73). Scientists would avoid reproducing exactly the observed sample in their image, since they feared prompting readers to mistake what was peculiar to this sample for a typical property of the object. As such, they would artistically render the object in its 'typical, ideal, characteristic' or 'average' form (Daston and Galison 2010, 69).

Atlas creators acting in line with the ideal of truth-to-nature thus deliberately presented their data in a manner that they did not believe Nature to have provided that data, in order to ensure their readership formed true beliefs. In the seventeenth and eighteenth century scientists could be upfront about this behaviour, since it was seen as good practice by the wider scientific community. Nowadays a truth seeking scientist may not be so forthright about correcting their data in this way, since it would be liable to be taken for misconduct. However, while we may not valorise such behaviour any more, the epistemological worry these authors were responding to remains. It is still possible to believe that some peculiarity of one's data or evidence would, if honestly reported, mislead one's community as to the truth of the matter. Truth seeking scholars may thus still be tempted to utter noble lies.

Hence both analysis of my model and consideration of the past behaviour of scientists shows that truth seeking scholars can be incentivised to utter noble lies, and thereby cause problems for motive modification anti-fraud policy. Further work in the motive modification tradition of anti-fraud research must show, if the tradition is to be viable, that something about present conditions ensures that truth-seeking scientists would not be tempted to utter noble lies.

6 Against individualism

The model itself suggests one candidate method of ensuring truth seekers are not tempted to commit noble lies; make them Self Confident. I therefore examine the relationship between my results and previous work in the field to get some sense of whether Self Confidence represents a desirable ideal for scientists. In particular, I compare my work to previous work by Kitcher. In apparent contrast with Du Bois' contentions, Kitcher argued that alethic puritans would make the scientific community as a whole less reliable when it comes to ascertaining the truth. The source of the apparent disagreement between Kitcher's results and Du Bois' contentions depends upon a distinction between communalist and individualist pure truth seekers. Kitcherian alethic puritan scholars are concerned only that they personally acquire true beliefs (Kitcher 1990, 14). Kitcher explicitly contrasts his model of pure truth seeking scientists with a model of scientists that are more altruistic in their desire for the community of scholars at large to arrive at true beliefs (Kitcher 1993, 344). Hence Kitcherian pure truth seeking, unlike DuBoisian pure truth seeking, would be naturally modeled by individualist, rather than communalist, truth seeking.

Author's personal copy

Kitcherian alethic puritans do not directly advance their goals by publishing. All they care for is the results of their own investigations; they benefit from reading other people's work, but do not gain from the community being aware of theirs. Correspondingly, in Kitcher's 1990 model his alethic puritans engage in no publication or direct information exchange.

However, if we assume that the esteem of the community is needed to retain access to the resources of the scientific community then a Kitcherian alethic puritan will behave in their publications as if they were a mixed credit/truth seeker. The only thing they stand to gain out of publication is having the scholarly community give them esteem; in effect, adding a concern for the communal consensus to their utility profile, making them into a mixed credit/truth seeking agent. But by the pessimistic theorem that need not make them an honest scholar, and by Du Bois' Conjecture that is likely to make them more dishonest than they would be if they were communally minded. Hence persuading scientists to be Kitcherian alethic puritans should not be any more likely to reduce the incentive to fraud than directly persuading them to be mixed credit/truth seekers, and would be worse than persuading them to be DuBoisian alethic puritans. Therefore my model can be seen to provide additional support to Kitcher's attack on epistemic individualism.

Relating this back to Self Confidence, a tension arises. The self confident agent is, in a certain sense, highly epistemically individualistic. In particular, they assign small ϵ credence to Nature favouring a signal other than that which they received even when they believe that literally everyone else in their field is due to report a result other than that which they obtained. This suggests strongly favouring the results of one's own work as compared to the results of communal endeavour. It is difficult to see how we might instill in scientists this kind of confidence in the importance and accuracy of their own work without also instilling in them a belief that their purpose as scientists is to discover the truth in or by their own work. If this difficulty is acknowledged, and in so far as a general moral of anti-individualism is the upshot of what is shared between Kitcher's work and my own, this mitigates against Self Confidence as an ideal for scientists' allocation of credences. This is, in turn, another blow to the motive modification tradition of anti-fraud research, given the previous results. Further work would be needed to see if motive modification could be effective under different suppositions about scientists' allocation of credences, or if despite appearances self confidence could be achieved without inducing the harmful effects of epistemic individualism.

On the subject of future work needed, it is worth drawing a more general moral for the social epistemology of science. In order to assess motive modification policy it has been necessary to consider how efforts that might be successful in curtailing fraud—in this case inducing Self Confidence combined with motive modification—would interact with broader social and epistemic goals one may have for the scientific enterprise. It cannot be assumed that the policy that would be most effective for minimising fraud is therefore worth endorsing; perhaps there would be trade offs which render such minimisation not worth the cost. My argument depends on the broad anti-individualism moral being taken as a reliable guide to the overall good of science; it is not worth compromising on this, I claim, in order to reduce the degree of fraud in science. But note that there are non-trivial assumptions involved

Author's personal copy

even in assuming that fraud, especially when takes the form of noble lies, is undesirable. Suppose truth-seeking scientists were very reliable in their judgements about what sort of signals from Nature are and are not misleading, as was supposed during the era when the ideal of Truth-To-Nature discussed by Daston and Gallison was endorsed by scientific atlas makers. Perhaps in that situation noble lies, as they are picked out in the model would be a net social epistemic good! I take the general moral to be that to test my claim that individualism is an undesirable trait to instill in scientists, and make similar arguments more precise, we need a workable model of the common good of science, suitable for use by social epistemologists. It is far beyond the scope of this paper to develop such a model, but I end this section by urging that such a model be developed.

7 Conclusion

There is a shared moral of this and Kitcher's work for social epistemology. As a slogan: scientists who are epistemically individualist hinder science considered as a communal enterprise. The role of noble lies highlights the degree to which epistemic individualism can be harmful. As can be seen from considering Active Honesty and No Insecurity together, even the communalist DuBoisian alethic puritan takes on board too much individualism if they are willing to let a desire to promote their own sense of what is true come before their communal duty of honest reporting. If No Insecurity were the best hope for motive modification anti-fraud research, it would risk running afoul of this anti epistemic individualism message. If the motive modification tradition is to survive, therefore, further work is needed.

In particular, such further work should focus on ways of reducing or eliminating the temptation to utter noble lies without inducing Self Confidence. One line of investigation that the model renders salient would be finding some way of reducing the extent to which science is socially stratified. After all, it is perceived decisiveness that the model suggests induces truth seekers to utter noble lies. Future work could hence explore the possibility and likely consequences of egalitarian social reorganisation in science.

Future work should also produce further generalisations of this model. As mentioned in Sect. 2, the model is designed to capture a specific sort of question scenario. It would therefore be worth generalising the sort of questions scholars can ask of Nature and the evidence they can receive from it. Further, it is implicitly assumed that one can publish no matter what result one reports. Whereas one source of temptation to fraud might be the idea that certain results (say, those reporting stronger effects) are more interesting to the community and therefore more likely to get published. It is an interesting question whether qualitatively similar results to those reported here would be found in a model of inquiry which explored that source of temptation to fraud. Likewise, the electoral structure of the model here prompts the thought that developing results concerning strategic voting may shed light on the phenomenon of fraud in science. There is hence potential for future formal work to expand on the modelling work begun in this paper.

Author's personal copy

Acknowledgments Thanks for helpful comments to Kevin Zollman, Carole Lee, Remco Heesen, Konstantin Genin, Jennifer Saul, Yuzuko Nakamura, Daniel Malinsky, Danielle Wenner, and anonymous reviewers for the Formal Epistemology Workshop and Philosophical Studies.

Appendix: Proofs

A publication market model consists of a question, consisting of a set of possible answers $\{a_1 \ldots a_n\}$, some scholars $\{v_1 \ldots v_m\}$, and Nature. Nature first selects an element of the question set, and then sends a signal to each element of the scholar set. Each signal is paired with one (and only one) answer, such that if scholar v_j receives signal s_i this raises the v_j credence in answer a_i without effecting v_j's judgements of relative credence between a_j and a_k (where $a_j \neq a_i$ and $a_k \neq a_i$). Call the element of the question set selected by Nature *the truth*, and the signal scholar v_j receives the *evidence proposition* received by scholar v_j. I call a specification of the question, the scholar set, and the decisions made by Nature a *configuration* of the model.

Each scholar chooses an answer to offer to their community. The election scholars take part in can be represented by a function from the profile of answers scholars offer to a winning answer. For ease of representation's sake I assume there is a unique winner, nothing will turn upon this assumption. Each scholar taking part in the election knows how the election function operates. Finally, the election function satisfies: if a_1 is victorious from profile $\mathcal{P}_\mathcal{A}$ and $\mathcal{P}_\mathcal{B}$ differs from $\mathcal{P}_\mathcal{A}$ only in the fact that all votes besides a_1 have equal or less votes in $\mathcal{P}_\mathcal{B}$ as in $\mathcal{P}_\mathcal{A}$ and a_1 has more votes in $\mathcal{P}_\mathcal{B}$ than in $\mathcal{P}_\mathcal{A}$, then a_1 is victorious in $\mathcal{P}_\mathcal{B}$.

Each scholar has a *utility type*, as per Sect. 2. I deal with three types in this paper. First, there is the *pure credit seeker*, who gains payout of 1 if the evidence proposition they announce is the winner of the election, and payout of zero otherwise. Second, there is the *pure truth seeker*, who gains payout of 1 if the winner of the election is the proposition Nature selected, and payout of 0 otherwise. Third, there is the *mixed credit/truth seeker*, who gains payout of 1 if they announce the proposition that wins the election and the winner of the election is the answer Nature selected, and 0 otherwise.

Each scholar has a *credence function*. This is a probability distribution over all elements of a relativised state space, defined as follows. A *relativised state space for scholar* v_j is a set of pairs $\langle p, a \rangle$ where p is the voting profile of every scholar except for v_j, and a is the answer chosen by Nature. So, for instance, if there are two scholars v_X and v_Y and two answers a_1 and a_2 the relativised state space for v_X would be: $\langle a_1^Y \rangle_1$, $\langle a_1^Y \rangle_2$, $\langle a_2^Y \rangle_1$, $\langle a_2^Y \rangle_2$. This would represent, respectively, the case where v_Y announces a_1 and Nature has favoured 1, where where v_Y announces a_1 and Nature has favoured 2, where v_Y announces a_2 and Nature has favoured 1, and where where v_Y announces a_2 and Nature has favoured 2.

Define a *response procedure set* as follows. Each element of v_i's relativised state space is an announcement profile specifying, first, which answer Nature has favoured and, second, how everybody but the scholar under consideration has voted. The response procedure set for a given element of the state space is constructed as

Author's personal copy

follows. Let c be an element of v_i's relativised state space. Let W_i be a function from v_i's relativised state space to the power-set of answers. $W_i(c)$ outputs the set of all answers that could win in c, depending on how v_i themselves announces. The output $W_i(c)$ is the response procedure set for c.

With the response procedure defined I can divide the relativised state space for a scholar into four types:

α: An element of v_i's relativised state space is an α element if and only if there is more than one element of its response procedure set.

β: An element of v_i's relativised state space is a β element if and only if its response procedure set is a singleton.

γ: An element of v_i's relativised state space is a γ element if and only if the answer initially chosen by Nature in that element of the state space is in their response procedure set.

δ: An element of v_i's relativised state space is an δ element if and only if the answer initially chosen by Nature in that element of the state space is not in their response procedure set.

Note that α/β and γ/δ are partitions of the state space. Finally, note that the following is true of response procedure sets:

Lemma 1 (Lemma 1) *If a_1 is in S's response procedure set for scholar v_1 then if v_1 votes for a_1 in S a_1 shall be victorious in the election over S.*

Suppose S is a β type set. Then no matter what v_1 votes for a_1 shall emerge victorious. Hence if v_1 votes for a_1 it shall be. Suppose S is an α type set. Since a_1 is in S's response procedure set there must be some vote v_1 could offer such that a_1 would win. Consider any such vote that isn't a_1, and call S filled in with that vote $S*$. Compare $S*$ to $S^\star$, which is S filled in with v_1's vote for a_1. Note that in $S*$ a_1 is victorious, and $S^\star$ is identical $S*$ with except that a_1 has one more vote. Hence if a_1 is victorious in $S*$ then it must also be in $S^\star$, and hence the lemma is proven.

I assume *scholars are expected utility maximisers*; scholars select an announcement to make which, given their beliefs about how likely they are to be in different elements of their relativised state space and their utility type, they expect to generate the highest return. I say that a scholar is *incentivised to fraud* if it would not be expected utility maximising to announce the evidence proposition that Nature sent them.

As mentioned in Sect. 2, I begin with a short demonstration that my representation of the mixed credit/truth seeker is superior to what might seem like a natural alternative. According to the alternative method of tempering the credit motive by way of truth, the scholar has a nuanced non-binary preference structure: receiving credit > no credit but truth victorious > other outcomes. That is to say, this scholar receives payout of 1 if the answer they vote for is victorious, $1 > r > 0$ if an answer they did not vote for but which Nature selects in this element of the state space is victorious, and 0 otherwise. Call this agent the *nuanced credit seeker*.

Lemma 2 *A nuanced credit seeker would vote for answer a if and only if a pure credit seeker in the same position would vote for a*

Author's personal copy

Two scholars are in the same position if they have the same relativised state space, the same credence over that state space, and received the same signal from Nature. Suppose nuanced credit seeker v^n was in the same position as pure credit seeker v^p. Consider the expected utility of v^p announcing a_i. It is equal to $\sum c_k$ for all elements k of v^p's relativised state space wherein c_k is a positive number and voting a_i attains payout of 1 in that element of the state space. Call the announcement that would maximise v^p's expected utility a_W. Consider the expected utility of v^n announcing a_i. It is equal to $\sum c_k + \sum rc_j$. This is the sum of all elements k of v^n's relativised state space wherein c_k is a positive number and voting a_i attains payout of 1 in that element of the state space, added to the sum of all elements j of v^n's relativised state space wherein c_j is a positive number and v^n voting a_i results in the community has voted for whatever the nuanced agents middle option is. Suppose a_W did not maximise v^n's expected utility, but some other answer a_F did. Note that since v^p and v^n face the same situation $\sum c_k^W \geq \sum c_k^F$— otherwise v^p would also prefer a_F. Hence it must be that $\sum rc_j^F > (\sum c_k^W - \sum c_k^F) + \sum rc_j^W$. Consider when rc_F will be earned. These are cases where the nuanced agents middle option but not top option is attained. That is to say, the community has selected the answer Nature selected, but v^n has not voted for it. If in such a case the community does not vote for a_W then $rc_j^W = rc_j^F$ would also have been earned by v^n in this scenario, since what Nature selects does not depend on what v^n voted for. Hence such cases cannot contribute to the left hand term being greater than the right in this inequality. However, if the community does vote for a_W then such cases contribute to $\sum c_k^W$ and since v^p and v^n have the same credences over states this case actually contributes more to the right hand side than the left hand side of the inequality. Hence whether or not the community votes for a_W the inequality cannot be satisfied. Hence a_W must also be v^n's expected utility maximising option.

Theorem 1 (Pessimistic theorem) *First, mixed credit/truth seekers can be incentivised to lie in scenarios where the pure credit seeker is not. Second, if a pure credit seeker is incentivised to lie in a scenario where a mixed credit/truth seeker is not, then in this scenario the scholar does not believe their vote will affect what proposition the scientific community comes to accept.*

Suppose without loss of generality that v_i received signal a_1. I consider the possible states v_i could believe themselves to be in, and how their behaviour would differ depending on whether they were a pure credit seeker or a mixed credit and truth seeking type.

α, γ elements of the state space. These are element where the scholar, v_i is decisive and their vote can bring about victory for the option they believe Nature to have favoured in this element of their relativised state space. Note that if a_1 is also the signal v_i believes Nature to favour in this state space, then v_i cannot be incentivised to dishonesty in this state space. If, however, a_1 is not the signal they believe Nature to favour, then the incentives of pure credit seekers and mixed credit/truth seekers can diverge. In particular, the following is true: if the signal v_i received is in the response procedure set for an $\alpha - \gamma$ element of the state space where the

Author's personal copy

signal favoured by Nature is not identical with the signal v_i received, a credit seeker cannot be incentivised to dishonesty while a mixed credit/truth seeker can. By lemma 1 a credit seeker could vote for the answer Nature signaled to them, and would expect to receive payout of 1 in such a scenario. Whereas, again by lemma 1, the mixed credit/truth seeker will only receive payout 1 if they vote for the option Nature favoured, which by hypothesis is not the answer Nature signaled to them. Hence in $\alpha - \gamma$ elements of the state space a pure credit seeker can never be incentivised to fraud while a mixed credit/truth seeker can.

β, γ elements of the state space. In such elements both pure credit seekers and mixed credit/truth seekers are incentivised to fraud just under the same conditions, namely just in case a_1 is not the sole element of the response procedure set.

δ elements of the state space. Note that in these elements a mixed credit/truth seeking scholar is in a state of despair: no matter what they vote for they believe they will get payout of 0. Hence they cannot be incentivised to fraud in any δ type state space. Whereas a pure credit seeker can be, depending on whether the signal Nature sent them is in their response procedure set.

To summarise: the mixed credit/truth seeker will not be incentivised to commit fraud, whereas the pure credit seeker might be, in scenarios where they believe they cannot bring the scientific community to accept the answer they believe to be true. However, in situations where, first, they believe the community is going to accept the truth however they vote, and, second, they do not think the evidence they received from Nature is representative of what is true, the pure credit seeker and mixed credit/truth seeker will be incentivised to lie at just the same times. What is more, if the two diverge and the scholar thinks they can bring the community to accept the truth *or* the answer their evidence supports depending on what they announce when they publish, then the mixed/credit truth seeker can actually be incentivised to dishonesty where the pure credit seeker would not be.

Now consider the behaviour of pure truth seekers.

Theorem 2 (Du Bois' conjecture) *A pure truth seeker is incentivised to lie on strictly fewer occasions than the mixed credit/truth seeker.*

Note that all of the above argument in theorem (1) would be identical for the pure credit seeker, with one exception. In β, γ elements of the state space the pure credit seeker takes themselves to be a guaranteed a payout of (1) no matter how they vote. Hence in such scenarios they cannot be incentivised to fraud, where the mixed credit/truth seeker would be. As such, the class of scenarios which, if believed to be most likely, would incentivise a pure truth seeker to dishonesty is a strictly proper subset of the class of scenarios which, if believed to be most likely, would incentivise a mixed credit/truth seeker to dishonesty.

The first behavioural posit mentioned in Sect. 2 is now explored.

Axiom 1 (*Cost of fraud*) Suppose scholar v_i's received signal s_h from Nature. Then for every element of the state space the utility of making any announcement $a_{k \neq h}$ is small $\epsilon_1 > 0$ less than it would otherwise be in that element of the state space given v_i's utility type.

Author's personal copy

This behavioural posit can be seen as representing the idea that there is ϵ_1 cost to fraud. This could be brought about if scholars attributed some small probability that they will be caught and punished for fraud. Note that this generates the following behavioural changes:

Theorem 3 (Active honesty) *If scholars obey Cost of Fraud, then the class of α, γ elements of the state space, where the signal the scholar received is in the response set, wherein a pure credit seeking scholar is incentivised to be honest is a super set of the class of α elements of the state space where a truth seeking scholar is incentivised to be honest.*

Say a scholar is incentivised to be honest wherein announcing the signal they received from Nature is their unique expected utility maximising option in a configuration. Consider an α, γ element of agent v_i's state space where the signal v_i received from Nature is in their response set. If it is, then a pure credit seeking v_i who obeys Cost of Fraud will, by Lemma 1, always be incentivised to honesty. This is because whatever v_i votes for will win, so they are guaranteed to be on the winning side, and Cost of Fraud gives v_i a preference for being on the winning side with their honest announcement. However, a pure or mixed truth seeking v_i may still fail to be incentivised to honesty. In particular, suppose in the α, γ element in question the signal they received from Nature is a_1 but they believe Nature to have favoured a separate answer, a_2, where a_2 is itself an element of the response procedure set. In this scenario the pure credit seeker would be strictly incentivised to honesty while the mixed or pure truth seeker would be strictly incentivised to dishonesty!

The second behavioural posit mentioned Sect. 2 is now explored. Consider the following behavioural posit.

Axiom 2 (*Self confidence*) Suppose scholar v_j received evidence proposition a_k from Nature. Let c_k be an arbitrary element of the state space wherein Nature favoured a_k and $c_{\neg k}$ be an arbitrary element of the state space wherein Nature favoured some answer other than a_k. Scholar v_j assigns small $\epsilon_2 \geq 0$ credence to any such $c_{\neg k}$.

Informally—scholars believe the results of their own research, and in particular if their research suggests Nature favours a_k then no matter what results they think their colleagues are going to report they still believe Nature favours a_k, assigning any other alternative such a small probability as to be swamped out in expected utility calculations.

Theorem 4 (No insecurity) *If scholars satisfy self-confidence and are pure or mixed credit/truth seekers then they would never be incentivised to commit fraud.*

As before, truth seeking scholars could never be incentivised to commit fraud in a δ type case, whereas a pure credit seeker still could. By definition and granting Self Confidence if the agent was in a γ type case they would believe the signal they received from Nature was in the response set. Hence whether they believed an α or β type element of their relativised state space was most likely, the truth-motivated scholar would not have incentive to lie if they were Self Confident.

Author's personal copy

References

Ben-Yehuda, N. (1986). Deviance in science—Towards a criminology of science. *British Journal of Criminology*, *26*(1), 1–27.

Broad, W., & Wade, N. (1983). *Betrayers of the truth*. New York: Simon & Schuster.

Bruner, J. P. (2013). Policing epistemic communities. *Episteme*, *10*(4), 403–416.

Casadevall, A., & Fang, F. C. (2012). Reforming science: Methodological and cultural reforms. *Infection and Immunity*, *80*(3), 891–896.

Cole, J. R., & Cole, S. (1973). *Social stratification in science*. Chicago: The University of Chicago Press.

Cole, S., & Cole, J. R. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review*, *32*(3), 377–390.

Dasgupta, P., & David, P. A. (1994). Towards a new economics of science. *Policy Research*, *23*(5), 487–521.

Daston, L., & Galison, P. (2010). *Objectivity*. New York: Zone Books.

Du Bois, W. E. B. (1898). The study of negro problems. *The Annals of the American Academy of Political and Social Science*, *11*(1), 1–23.

Fanelli, D., Costas, R., & Larivière, V. (2015). Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity. *PLoS ONE*, *10*(6), 1–18.

Hintikka, J. (2007). *Socratic epistemology*. Cambridge: Cambridge University Press.

Kitcher, P. (1990). The division of cognitive labour. *The Journal of Philosophy*, *87*(1), 5–22.

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford: Oxford Univeristy Press.

Kuhn, T. (1996). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kummerfeld, E., & Zollman, K. J. (2015). Conservativism and the scientific state of nature. *British Journal for the Philosophy of Science*. doi:10.1093/bjps/axv013.

Latour, B., & Woolgar, S. (1986). *Laboratory life*. Princeton: Princeton University Press.

Merton, R. K. (1968). Behavior patterns of scientists. *American Scientist*, *57*(1), 1–23.

Merton, R. K. (1973). *The sociology of science*. Chicago: University of Chicago Press.

Muldoon, R. (2013). Diversity and the division of cognitive labor. *Philosophy Compass*, *8*(2), 117–125.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia II: Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615–631.

Solomon, M. (2001). *Social empricism*. Cambridge, MA: The MIT Press.

Sovacol, B. K. (2008). Exploring scientific misconduct: Isolated individuals, impure institutions, or the inevitable idiom of modern science? *Journal of Bioethical Inquiry*, *5*(4), 271–282.

Strevens, M. (2003). The role of the priority rule in science. *The Journal of Philosophy*, *100*(2), 55–79.

Strevens, M. (2006). The role of the Matthew effect in science. *Studies in History and Philosophy of Science Part A*, *37*(2), 159–170.

Zollman, K. J. S. (2010). The epistemic benefits of transient diversity. *Erkenntnis*, *72*(1), 17–35.

Zuckerman, H. (1988). The sociology of science. In N. J. Smelser (Ed.), *Handbook of sociology* (pp. 511–574). Thousand Oaks: Sage.