The identification and prevention of bad practices and malpractices in science.

Commentary on Hanne Andersen's
"Epistemic dependence in contemporary science: Practices and malpractices"

Watch out. Many typos were introduced in the proofs by the editors, so I do not have a clean edited version of this article, with all proper references to Andersen's article and corrections.

Please use the published version when citing or quoting.

Imbert, Cyrille. « The Identification and Prevention of Bad Practices and Malpractices in Science: "Commentary onEpistemic dependence in contemporary science: Practices and malpractices" by Hanne Andersen ». édité par Léna Soler, Sjoerd Zwart, Michael Lynch, et Vincent Israel-Jost, Routledge, 174-187. Routledge Studies in the Philosophy of Science. London, 2014.

According to Hanne Andersen, "an analysis of <malpractices> goes beyond research ethics and includes important epistemological aspects" (p.1). Her purpose is to point at a new area for philosophy of science in practice, which she does by highlighting different epistemological issues about malpractices and showing how documenting them in a precise way is beneficial to their solution. She articulates in particular two questions, namely the issue of the identification of bad practices and malpractices, and the ways of preventing the latter from happening. I shall discuss how Andersen contributes to these issues, and make additional suggestions.

Before going further, I first want to clarify a few points.

a) Scientists can fail to match scientific standards, as scientists, in various ways. In what follows, I shall focus upon ways in which scientists fail to follow scientific standards in their *research* practices, which leaves aside other important circumstances in which they may scientifically misbehave. For example, casting doubt about scientific issues by challenging scientific evidence on scientifically non-accepted grounds in fields in which one is not an expert, as Frederick Seitz or Fred Singer did about issues like global warming, is an example of using one's scientific reputation to smuggle in pseudo-expertise, a clear scientific misbehavior (Conway and Oreskes, 2010), but not a research malpractice.

b) Something that is presented as a good research practice can fail to be so for various reasons. To mention just a few, a practice may not contribute to a target result in the way her author claims that it does; or it may have been carried out by its author without the scientific vigilance or care expected by her community; or, it may not be a token of a scientific research practice at all – typically, forging data is not an example of a bad scientific practice, it is simply no scientific practice at all. I shall follow Andersen in describing as malpractices all such cases, when the failure occurs either deliberately or by negligence. The difference between the two is that, in the latter case, the author does not intentionally want to perform a bad practices what happens but consciously fails to follow an accepted standard of scientific rigor, while knowing it may adulterate her practice and results.

c) One may wonder whether the above working definitions are a sufficient basis for tackling issues about malpractices. From a logical point of view, analyzing intentional bad practices, requires possessing a sound characterization of what good/bad practices are, and therefore having a sound definition of scientific practices themselves, a question that is still being investigated. Further, the dichotomy between good and bad may be too coarse. For example, a trichotomy bad/acceptable/good may catch more precisely the epistemic stances of scientists towards actual practices. Typically, a reviewer may deem a practice acceptable for publication but may not wish to rely on the corresponding paper for her research. Finally, in a research community, it is unlikely that there is a consensus about where the boundaries between good, bad and acceptable practices lie: while shared hypotheses and methods provide a common basis for scientific discussions, they do not determine every single aspect of practices and this provides room for disagreement about the validity of this or that aspect of a practice. But do we really need to solve all these issues to start analyzing malpractices? Probably not. Should we provide a principled analysis that would be sufficient for understanding what would happen if, for each scientific practice, we asked scientists to discuss whether it is a case of good, acceptable, bad or non scientific practice, we would have to answer all the above problems; but, as Andersen's paper illustrates it, it is possible to fuel a valuable discussion about malpractices by focusing on uncontroversial cases and this is enough for the present purposes.

I. Identifying bad practices and malpractices

While one of the declared topics of Andersen's paper is the identification of scientific malpractices, the focus in her paper deals mainly with the "calibration of trust [and distrust]" between scientists. Typically, she refers to direct calibration (when a scientist directly assesses the result of another scientist because it belongs to a field that overlaps with her own field of research) as a way to calibrate scientists, not their practices. Of course, the calibration of practices is often not an available option for scientists and trust has to come into the play (Hardwig, 1991). But, the direct assessment of the value of scientific practices is an important part of the identification of bad practices, and it should be analyzed in details, if only to understand how delicate this assessment is, if it is to be carried it out properly, and why bad practices and malpractices are difficult to erase in science.

1.The identification of bad practices and malpractices, or the new trickier version of the demarcation problem

Why are not bad practices more easily identified in science? I shall first hint at some principled reasons why this is so, and explain why the picture of science that emerges from the practice turn gives insightful, if not explicit, clues regarding this issue.

This identification question is germane to the problem of demarcation between science and pseudo-science, which was seen as central by logical empiricists[1]. According to this tradition, focusing on features of scientific statements (theories, axioms, observational statements, etc.) and their treatment by scientific reasoning is a sound and fruitful way to analyze scientific activity. Then, the question of the demarcation between scientific and non scientific practices or good and bad ones (resp. research programs, agents, particular inquiries or any relevant

[1] Logical empiricists claim that they want to distinguish between metaphysics and science. Because they provide a solution based on a characterization of scientific statements, they de facto answer the more general problem of the demarcation between science and non-science (for more details, see Hansson, 2012).

scientific item) can still be raised, but it should be seen either as having a solution that can be derived from an analysis of more central (that is, linguistic) features of science, or as "merely" related to the pragmatics of science.

On the contrary, once one adopts a perspective in which scientific practices are seen as the core of science, and the right level of analysis to tackle traditional questions like the meaning of scientific terms or the nature and possibility of scientific progress (see in particular (Kitcher, 1993)), it becomes of crucial importance that one understand better how good scientific practices can be distinguished from bad or unscientific ones and how demarcation issues are solved in practice. The worry, as I shall now argue, is that this version of the demarcation problem is much trickier to solve.

There is in general a gap between the possession of a sound, clear, and precise definition of an X and the ability to recognize particular instances of this X when they are met: knowing what prime numbers or proofs are is one thing, identifying instances of them is another. In the case of science, the identification problem requires studying the gap between the predicates 'being a scientific (resp. acceptable scientific, good scientific, bad scientific) item' and 'being identifiable as a scientific (resp. acceptable scientific, good scientific, bad scientific) item'. And there is of course another gap between 'being identifiable as a scientific item' and 'being identifiable as a scientific item with limited resources and on the basis of limited information'. As I shall now argue, there are various reasons why these gaps should be wide in the case of scientific practices.

First, even if science and scientific activity could be fruitfully described by merely analyzing scientific statements and calculus of scientific reasoning, the identification problem would remain. As is well known, whether a sentence belongs to a language can be undecidable or require computationally untractable procedures to be solved, even if there does sometimes exist tractable procedures to answer such questions. So even if the demarcation problem boiled down to questions like "is this sentence a consequence of this theory T" and theories were axiomatized, the question of the identification of *bona fide* scientific items would be difficult in practice.

Second, because scientific practices are much more complex and multi-dimensional entities than meaningful statements of theories or proofs, the difficulty to identify the latter should be a lower bound to the difficulty to identify the former[2].

Third, in many cases, there is often no way to give a complete access to practices, which may make it difficult to settle in a transparent, explicit and uncontroversial way potential debates about their validity. While scientists do share elements of practices – hence the notion of "consensus scientific practices" developed by Kitcher (1993) – many of these are not explicit (see introduction, p.XX, see also (Collins, 2010), (Soler, 2011)). This may be the case for elements like typical experiments, ways of identifying authorities, scientific values, typical instruments and familiarity to use them or elements, like methodological principles or rules of thumbs, which, though sometimes conveyed in natural language, do not have a precise semantics. As a consequence, even what is shared cannot be made common knowledge (by use of public announcements) and it can never be completely uncontroversial that an individual practice is an adequate token of an un-explicitly agreed upon practices. And of course, individual practices include in addition more specific instruments, particular

[2] If a subproblem A' of a problem A has difficulty K, the more general problem A cannot have difficulty less than K. Among other things, scientific practices include linguistic components and judging the quality of a practice often requires, among other things, judging the quality or validity of a linguistic, mathematical, or computational item.

experiments, original methods, runs of simulations and large sets of data, etc., all of which cannot be completely presented in articles. As a consequence, judgments about the value of practices often need to be made on the basis of irreducibly incomplete information, and this is more room for misidentification of valid practices. Importantly, this problem is not simply one of moral integrity or deliberately partial reports made by scientists about their individual practices. When honest scientists unconsciously fail to carry out good scientific practices, their peers may not detect their failures because the reports do not and cannot include all the relevant scientific information about potential causes of failures, and experts may fail to ask about all aspects that may conceal hidden unexpected problems.

Finally, as pointed above, there can be vagueness remaining in some aspects of a research progam, as well as disagreement between sub-communities or individuals about which practices should be considered as good ones, all of which can be potential sources of troubles for the uncontroversial identification of bad practices. And of course, the more complex an item is, the more it is likely that we have partial disagreement or knowledge about its nature and potential troubles in consensually identifying accepted instances of it, and scientific practices are complex multi-dimensional entities for sure.

Overall, the difficulty to solve the identification problem means that, in many cases, even experts like reviewers, have no infallible procedure for identifying and discarding results issued from bad practices. In other words, we have some principled reasons explaining why the dream of a perfectly and infallibly checkable science definitely should be seen as a utopia. And it is no surprise that, as Andersen points it out, the identification of malpractices is a difficult issue and that there remains a potentially large grey zone in science – that is, published results that are the product of poor research due to "sloppiness", "incompetence" (p.1) or "so-called honest error" (p.6), given that – by definition – what belongs to this grey zone is not clearly known.

The implications of the above general negative conclusion should be drawn with great care. It should not for example be seen as implying that the situation is always desperate and that scientists have no good ways of checking individual results. To use an analogy, a problem can be in general undecidable but be composed for a great part of decidable problems; and difficult problems can have easy subproblems or easy approximate solutions. In the same way, saying that the identification problem cannot always be exactly solved is not informational about how often it can be solved, how much there exist approximate methods that are often successful, etc. From this perspective, a reasoned analysis explaining how much identifying bad practices is difficult, depending on the type of practices involved, is still to be made.

As things are, it is not surprising that the examples of malpractices described by Andersen are drawn from experimental parts of the natural sciences. But how much should one extrapolate? While one may agree that the analysis of science in terms of practice is a general one and includes the formal sciences – Kitcher's seminal analysis (1983) was indeed about mathematics (see also (Giardino et alii, 2012) and (Mancosu, 2012)), the philosophy of scientific practices has devoted much attention to experimental science and it remains to be seen how much the practice turn differently or similarly impacts our understanding of the various sciences regarding each question, and this one in particular: in other words, a comparative analysis of what it takes to identify bad practices and malpractices in the various sciences, and which aspects of practices are responsible for these differences, would be most welcome.

To wrap up, once it is acknowledged that the identification problem cannot *always* be solved with total reliability, there remains to be analyzed how the *direct* epistemic identification of individual bad practices does work, that is, when, how much, how and with what reliability it is possible to directly assess the validity and quality of scientific practices by analyzing information and clues about their nature and content. And this philosophical agenda is compatible with the claim that the indirect calibration of practices, which is made by using clues related to the external circumstances in which they were carried out (like the reputation of their authors or the institution in which they were developed) also plays an important role in science.

Because evaluation based on the description of practices is difficult, it is not surprising that scientists also use indirect coarser strategies such as calibrating the practitioners themselves, all the more since calibrating individuals on the basis of external indicators can be less costly than analyzing in details their practices. From this point of view, the difficulty to solve the identification problem is one more reason to follow Hardwig and say that trust is an essential ingredient of science – perhaps "even more basic than empirical data or argument" (Hardwig, 1991, 694): because external agents cannot check the validity of practices, we have in part to trust authors, both epistemically and morally, for science to be possible. But indirect strategies cannot be all there is to the evaluation of practices and they need to be in part fed in somewhere in the process, by direct partly reliable epistemic evaluations of practices. Further, as already pointed out, authors can be honest and competent but fail to develop fully satisfactory practices – and there is then the need that external evaluators help them find out when something is wrong and pin down why – and for this reason, trust towards authors cannot be the sole answer. Reviewing is often described as a practice of selection. It surely in part is, at least in a first stage; but it is also a practice of melioration and the finally accepted papers are usually significantly better than they would have been, had not the reviewers requested revisions by calibrating practices and results. So even if the problem of identification of good and bad practices has no general solution, there are ways to partly solve it, and the possibility of reviewing, as a melioration activity, is an evidence of this.

Overall, given the difficulty of the identification problem, it is not surprising that both strategies, direct and indirect, are sometimes jointly used. There is clearly the need to analyze how both strategies work, given that the study of the direct evaluation of practices (by reviewers and peers in general) probably needs to be rooted in case studies and is work for the philosopher of practices. Another question is to analyze which balance of direct and indirect strategies are acceptable, if not optimal, if science is to progress reasonably, and this seems to be work for social epistemogists[3].

2. Calibrating good/bad/mal practices and practitioners: what relations?

Let us now turn to the calibration of scientific practices via the calibration of practitioners, and the assessment of when they should be deemed trustworthy, which is how Andersen tackles this issue. Believing results without having calibrated the corresponding practices means that one becomes epistemically dependent on its author, who is the warrant of their validity and is trusted. Andersen follows Hardwig's analysis of trust, which says that it is based on beliefs about the epistemic and moral character of the author. Accordingly, if B is

[3] Given that there are various ways of using testimonies and external clues in general, there are many ways of indirectly calibrating authors and practices. Finding which ones are most efficient is another object of inquiry (see in particular Mayo-Wilson, XXXX).

trusted by A, B will be believed to be knowledgeable and truthful (given that it is unlikely that someone is trustworthy and untruthful, that is, reliable by accident); but "untrustworthy scientist may either be untruthful or unknowledgeable (or both)" and, to calibrate untrustworthiness, there may be the need to distinguish between the moral and the epistemic component of a scientist's trustworthiness" and "assess [her] moral and epistemic character separately" (p.4).

While it is commonly accepted that trust, and trust in moral integrity, is involved in science (Rennie, 1997), there is still work to be done to delineate how much and when it does play a role. I shall in the rest of this section content to analyze when the calibration of practitioners and their moral character in particular seems – or not – to be required.

First, as partly discussed at the end of last section, even if one is not a reductionist believing that all legitimate trust in scientific agents is rooted in the evidence-based assessment of how knowledgeable (and moral) agents are, calibration of agents cannot be the whole story, and the calibration of the very practices is needed somewhere in the process, if only to give sometimes reliable clues about the reliability of agents. A good description of how trust is built should disentangle the role of the various components that contribute to it. As an aside, even if the calibration of practitioners were exclusively grounded on the results of the calibrations of practices, it would still make sense to acknowledge that it does play an independent role in science. Indeed the calibration of practices through the assessment of results and how they were reached (what reviewers typically do) and the calibration of practitioners do not arise in the same epistemic circumstances (see the third point below); further, how exactly the results of the calibration of practices (typically, acceptance in this or that journal) should be used in order to build a picture of the reliability and trustworthiness of scientists is an independent question.

Second, *as researchers*, scientists are first and foremost interested in reliability. If it is shown that they do not always need to, why should they get engaged in moral evaluations or beliefs? Indeed, belief in a degree of reliability of an agent does not commit to any particular belief about the morality of this agent. For example, if I believe that an author A has a 0.9 overall degree of reliability on the ground of indicators like her publication records (that is, ways that do not require the direct moral calibration of A), my belief is compatible with different beliefs about her moral and epistemic character (e.g. having a 0.9 degree of integrity and 1 degree of competence, or vice versa). Therefore, I do not need to be committed to any particular belief about her moral or epistemic character– even if in the process of acceptance of papers, primary evaluators, like reviewers, may have entertained beliefs about her moral character (especially if data were reported), and I need myself to calibrate my trust towards these unknown evaluators or indicators. So, trust in the results of an author A at best implies having an implicit indefinite belief about her moral and epistemic character that is compatible with one's precise degree of trust, and no precise commitment is required. In other words, Hardwig's analysis of the implication of trust in terms of beliefs about the truthfulness of authors need not always accurately describe the actual beliefs of scientists about the authors that they epistemically depend upon.

Third, if trust, on the one hand, and belief in moral and epistemic character, on the other, are related, then assessing the epistemic and moral character of A *can* be a way to assess the reliability of A (and vice versa). For epistemology in practice, the question is then "which ways to reliability and trust are usually taken?" There clearly seems to be cases in which the way that goes through the calibration of moral character will not be taken and no explicit belief about the morality of authors will be developed. For example, clues about reliability can derive from external indicators like the scientific records of the author, the reliability of the journals she published in, the credit of her co-authors, etc. All this may be sufficient evidence for believing it is rational to trust a result and its author – and no explicit moral trust, let alone direct calibration of moral character, are required.

This does not however imply that moral calibration plays no role in the scientific processes that lead to trust by an agent towards a particular result. Potential reliability indicators (reputation of scientists, quality of a journal, etc.) also need to be trusted and calibrated by agents and this may be done on different grounds (information about the reliability of past results, moral character of editors, trust in the judgment of esteemed peers, etc). Primary procedures of evaluations of new results – typically reviewing – may also partly require trust in the moral character of authors (see below). Still, the existence of moral trust in the evaluation process differs in various ways from the existence and use of explicit moral trust towards authors by scientific users. First, even if beliefs in the moral character of authors play a role in the evaluation process, this needs not propagate downstream. Indeed, belief does not seem to be a transitive relation (if I believe that a journal editor is reliable, the journal editor believes that the referee is reliable and the referee believes that the author is morally honest, it does not follow that I believe that the author is honest). Second, even if reviewers have to morally trust authors, the dependence on moral trust may vanish later in the process, when the reliability of published results is further checked by the community, and the result becomes well-entrenched (or not). One may finally note that it may be safer to rely on beliefs about the moral character of editors or colleagues regarding a journal than on beliefs about the moral character of authors regarding their own results.

Fourth, it is probably the case that, in practice, calibration is performed differently in different scientific contexts and that calibration of bad and good practices (or authors) work differently (even if 0.8 trust is 0.2 distrust). Here are examples to illustrate this point and show how it could be developed.

i) Scientific judgments between peers involved in research. First, it is likely that only calibration of "good enough" practices will usually be completed in research contexts. Arguably, scientists are not after a detailed evaluation of the practices of their peers but only want some good reasons for trusting a subset of very reliable results that pass a chosen threshold, and the distinction that matters is between results under or above this threshold. As soon as it becomes obvious that a result will not pass this threshold, it needs not be calibrated any further – precise assessments take time – since the reasons of this reliability failure do not matter for research – unless perhaps the result is of crucial importance and one is compelled to precisely assess its value on exclusively scientific grounds. To take one of Andersen's examples, once Nobel Laureate Peter Medawar was convinced that there was something fishy about Summerlin's work, he did not bother to push forward the investigation and determine whether this was a case of incompetence or dishonesty.

In other words, scientists are like diggers that try to find gold nuggets in a mine and procedures for finding these nuggets need not be similar with (never actually used) potential procedures for classifying all clearly-non-golden nuggets. This analysis is coherent with the fact that only a small fraction of the literature is cited and it is not unlikely that this fraction complies on average with highest epistemic standards. So, the argument goes, as far as the advancement of research is concerned, no calibration of moral character is needed for the most dubious results, for which moral calibration would be most needed.

ii) Moral trust without moral calibration as a default rule. In some cases, typically in experimental research, having extremely high trust on purely epistemic evidence may not be possible and authors may have to be morally trusted regarding the uncheckable aspects of their work. This does not however imply that moral calibration is then performed. Blind moral trust for what cannot be checked, may have to be the default rule, as acknowledged by some scientists (Rennie, 1997, 579). Similarly, when they accept a paper after a thorough epistemic scrutiny, reviewers are more epistemically vulnerable than scientists using the literature, since they cannot benefit from the expertise of other members of the community, which is present once a result has been published and discussed and the result has become entrenched. Reviewers are on their own and, for this reason, they may have to trust to a greater extent the epistemic and moral character of authors. This does not however imply that reviewers develop a specific activity of assessing the morality of the author – again, how would they do? It can be argued that they simply try to eliminate bad practices and select good ones on the basis of available epistemic evidence – and defectors that do no play the game honestly regarding uncheckable aspects of their work and are afterwards identified will be given a tit for a tat by no longer being trusted[4].

Actually, it is not even sure that reviewers are committed to believe that submitting authors, whose papers are accepted, are honest, since they may be simply described as doing *as if* authors were honest, given that honesty and trust are the condition of possibility of scientific activity, and are therefore the rules of the game. So they may be described as saying something like this: "to the extent that it can be checked, the content of the paper is worthwhile and the amount of requestable epistemic evidence that has been provided is proportionate to the importance and novelty of the paper – given that any evidence cannot be requested, in particular in the case of simulations or experiments; so I have the conditional belief that if the author has been honest and conformed the ethos of scientists, this is a good result".

iii) Fraud detection. Once malpractices are publicly suspected for a result, it clearly becomes an important issue to determine whether an author should be convicted of malpractice. Inquiries aimed at assessing the morality of authors may then have to launched, with distinct – heavier, institutional, more collective – procedures than what happens in the usual processes of peer calibration, and the cases presented by Andersen nicely fits this description.

iv) Collaboration. As clearly highlighted by Andersen, collaborators are epistemically vulnerable towards their co-authors. Because being engaged with fraudulent co-authors is risky, and co-authors are sometimes in a position to have additional clues about the honesty of their colleagues, calibration of malpractices is more likely to take place.

Overall, it is dubious that beliefs about the moral character of authors and moral calibration is explicitly involved or at play in all cases of epistemic dependence, even if potential beliefs in the moral character of author may be an implicit consequence of an actual belief – but as is well-known, having the actual belief A needs not imply actually believing all the consequences of A. So, in which scientific circumstances moral beliefs and moral calibration are actually important would have to be investigated in more details.

This being said, even if one subscribes to this mitigated skepticism about the importance of moral beliefs and calibration in scientific activity, and even if one believes that identifying malpractices is much more difficult than identifying bad practices, and that researchers do not frequently engage in such activity, there is room to agree with Andersen that malpractices are an important object of inquiry for philosophers of science and epistemologists. Indeed, even if

[4] See (Blais, 1987) for applications of the tit for tat strategy to knowledge contexts.

bad practices and malpractices are hard to identify, understanding what they are, how they occur, why, etc. cannot but be useful to make them less frequent. A parallel can be drawn with the detection of driving infractions. While identifying all cases, intentional or not, of dangerous driving infractions is hardly possible, understanding which mechanisms contribute to generate them can be instrumental in reducing their number. So it is clearly important that epistemologists investigate which scientific policies (like editorial policies about co-authorship) can be adopted to reduce malpractices.

II. How to fight against bad practices and malpractices – and when?

Scientific bad practices may be a threat both for science and society, and their detection is by no means easy. What can be done, then, to make them less frequent in science? Because co-authors have information that reviewers and readers do not of have, they are in a better position to be aware or suspicious of scientific faulty practices. Accordingly, one of Andersen's suggestions for improving detection is to make co-authors partly responsible for the fault of their colleagues and thereby compel them to be whistleblowers.

1. Differentiating questions about efficiency

Before discussing Andersen's suggestion and its justification, let us analyze more sharply the general issue of fighting against malpractices. Bad practices are faulty actions, which can have detrimental consequences. When trying to eradicate them, one must watch out to assess all the effects of the policies that one may want to apply. Accordingly, it is important to distinguish between the following questions.

- P1. Efficiency of prevention problem. Which policies can be adopted to keep bad practices and malpractices as low as possible?
- P2. Scientific efficiency problem. Which policies are most scientifically efficient (and, in particular, are the policies that keep bad and malpractices low scientifically efficient, once taken into account all their epistemic and scientific effects)?
- P3. Social efficiency problem. Which policies are socially efficient, once taken into account both social and scientific advantages and drawbacks?

Let us be more explicit. It is hardly controversial that, everything being equal, if there are ways to decrease the importance of bad practices and malpractices in science, they are welcome and it is important to identify such possible ways and answer P1. The distribution of responsibility among co-authors is such a potential repellant against malpractices. Still, in trying to fight against malpractices, one must watch out not to make scientific activities, and collaborative practices in particular, too risky and thereby hamper the development of science, which may crucially require collaborations. So it is important to analyze how beneficial such potential policies against bad practices are for science in general, which means answering P2. Finally, given that bad practices can also have dramatic consequences on society, we should not content ourselves balancing scientific advantages and drawbacks only. Some policies may be globally detrimental for science, because they slow down its dynamics, but beneficial for society, because they filter out some bad practices, even if a few of them, that may have devastating social consequences. Answering P3 therefore requires an epistemological and social analysis of the effects of such policies and weighing issues like how much it is acceptable to tolerate minor scientific risks of not detecting bad scientific results if they imply major social risks for society, which can be the case when health or environmental issues are concerned.

2. Malpractices and the distribution of responsibility within collaborative works

There are various ways of fighting against bad practices and malpractices. Some are targeted at individual authors and consist in the development of policies aimed at preventing and detecting bad practices or malpractices at the individual level. Other can be organized at the institutional level of scientific research, such as policies controlling the funding of research and its transparency or potential conflicts of interests. Andersen focuses on the individual level when she analyzes how, in the context of collective works, responsibility for individual malpractices should be shared by co-authors, based on what she calls a relational account of calibration, in which "the strength of the calibration is dependent on the epistemic character of the researcher performing the calibration" (p.9). She gives as a potential justification for shared responsibility the fact that co-authors have access to technical details that other agents, such as referees, editors or scientific users do not have: with greater knowledge and the possibility of making valuable calibrations comes greater responsibility. One may wonder however whether this responsibility is rooted in the *actual possession* of knowledge of a crime or *the possibility* to have easier access to evidence and a correlative epistemic duty of calibrating co-authors. This latter option seems more appropriate to describe the Woo Suk Hwang case and the position of Gerhard Schatten, the senior researcher who, given his role and knowledge, was in an epistemic position to be suspicious about the possibility for his junior co-author to have really carried out the research he pretended he had, and could gather evidence to prove the fault. Andersen also points at that "collaborative research requires trust", which must "be balanced with a responsibility to ensure the veracity of all results" (p.10). Here again, the claim is suggestive: scientists who gave their trust should be epistemically accountable for giving it. But then, are co-authors partly responsible for the malpractice or simply for their epistemic failure to detect it – a much more benign fault, given that, the lesser the fault, the lesser the punishment, and the less efficient the prevention?

Andersen, finally grounds the potential responsibility of co-authors in an analysis of the epistemic duties of members of scientific groups, who, "participating in a shared cooperative activity of delivering – and being able to defend – a new and interesting result *p*" (p.10), have the duty of being "epistemically responsive" and having "meshing subplans", which "must include preparing for serious critical inquiries" (p.10). But here again, there is a difference between being responsible for failing to organize an inquiry – a benign epistemic fault – and the actual responsibility for malpractices to which co-authors did not actively contribute to. She concludes that the very author of a malpractice can be morally blamed but that co-authors can only be epistemically blamed "in exactly the same way as had the data been caused by a defective instrument instead of a defective collaborator" (p.10). While this is a suggestive distinction, it does not help assessing how much the responsibility should be shared, since there are indeed cases (like parents for children or ministers for their administration) in which it is possible to be responsible, and even liable, for a fault that one did not wanted.

How to attribute responsibility and credit is a difficult question. The directions indicated by Andersen seem to be valuable ones but they raise significant conceptual and philosophical questions about the nature of agency – something that responsibility is often rooted in –, the various responsibilities and accountabilities involved – moral, epistemic, scientific, legal – and how potential punishments should be tied to how responsibility distributes. Since scientific practices are actions having their specificities, it surely requires an input from scholars studying science. My final suggestion is that the treatment and clarification of these questions may also benefit from, if not require, taking into account the existing rich debates in philosophy of law and action, in particular about issues like collective responsibility and cases in which there may be responsibility "with non contributory fault" (Feinberg, 1968, 681). Malpractices are, after all, faulty activities among others and they are sometimes embedded in larger criminal activities (e.g., when the malpractice is deliberately aimed at favoring some industrial interest) by making apparently legal an activity that would not be permitted, had the right results about the corresponding scientific questions been made public. As indicated by the ongoing debates about authorship, responsibility and accountability in scientific journals

(Rennie, 2006, Egert, 2011), these are questions that are in present need of treatment for the development of a healthier science, and philosophers can contribute to shaping the forms that the authorial and editorial scientific practices should take in the future.

References. (Only references not quoted by Andersen are mentioned). 238 words.

Blais, Michael, (1987) "Epistemic "Tit for Tat", *The Journal of Philosophy*, Vol. 84, No. 7 (Jul., 1987), pp. 363-375.

Collins, Harry, (2010), Tacit and Explicit Knowledge, Chicago: University of Chicago Press.

Conway, Erik and Oreskes, Naomi, (2010), *Merchants of doubt: how a handful of scientists obscured the truth on issues from tobacco smoke to global warming*, London : Bloomsbury Press.

Eggert, Lucas D., (2011) "Best practices for allocating appropriate credit and responsibility to authors of multi-authored articles", *Frontiers in Psychology*, volume 2, article 196

Feinberg, Joel, (1968) "Collective Responsibility," *Journal of Philosophy*, 65: 674–688.

Giardino, Valeria, Moktefi, Amirouche, Mols, Sandra and Van Bendegem, Jean Paul, (2012) *From Practice to Results in Logic and Mathematics*, special issue of *Philosophia Scientiae*, 16 (1).

Hansson, Sven Ove, (2012), "Science and Pseudo-Science", The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/pseudo-science/>

Kitcher, Philip, (1983), *The Nature of Mathematical Knowledg*e, Oxford University Press

Kitcher, Philip, (1993), *The advancement of science: science without legend, objectivity without illusions*, Oxford University Press

Mancosu, Paolo (2012), *The Philosophy of Mathematical Practice*, Oxford University Press.

Mayo-Wilson, Conor, (to appear XXX), "Reliability of Testimonial Norms in Scientific Communities", *Synthese.*

Rennie, D., Yank, V., & Emanuel, L. (1997). When Authorship Fails: A Proposal to Make Contributors Accountable. *JAMA, 278,* 579-585.

Soler, Lena, (2011), "Tacit aspects of experimental practices: analytical tools and epistemological consequences", *European Journal of Philosophy of Science*, 1:393–433

Sox, Harold C. and Rennie , Drummond, (2006) "Research Misconduct, Retraction, and Cleansing the Medical Literature: Lessons from the Poehlman Case", *Annal of Internal Medicine*, 144:609-613.