IUPAP Workshop on Scientific Misconduct
and the Role of Physics Journals in its Investigation and Prevention
13/14 October 2003, Institute of Physics, 76 Portland Place, London

13 October:

Roger Elliott welcomed the group and began the meeting. As incoming chair of the IUPAP Working Group on Communications in Physics, he mentioned that the group has so far considered issues such as linking/mirroring, intellectual property, communication in electronically remote areas, eprint archives, and peer review. He observed that closer scrutiny resulting from the recent data fabrication cases in physics has made it clear that lower level fraud in the field is more widespread than had previously been assumed. He invited Martin Blume, outgoing group Chair, to open the meeting.

Misconduct in scientific publication is not new, Blume said, but public interest in it is. Cases in biomedicine were first in the public eye, and they struck a nerve. Physics appeared relatively exempt from these problems, but that sadly is no longer the case. A thoughtful, complete and visible response is essential to regain the public trust. He hoped that the workshop discussions would form a basis for a document that IUPAP would be willing to adopt, outlining the responsibilities of the various parties in the physics research and publications enterprise. He further introduced the subject with a series of slides covering the various types of misconduct and which party (journals, institutions, agencies) was responsible for investigation. He emphasized the importance of preserving confidentiality during investigation, since careers are often at stake.

The first topical session of the meeting concerned **plagiarism**, which is probably the most common type of misconduct. Session Chair Erick Weinberg noted that plagiarism always consists of copying ideas and very often of copying text as well. He described several cases, such as the paper by the authors Rajput and Joshi (http://www.geocities.com/physics_plagiarism/) which was essentially a copy of a paper by Renata Kallosh. He also cited another example of a paper on ArXiv hep-th in which the author had essentially reversed his initial conclusions in response to a critical posting, with the revised version appearing only hours after the critical article had been posted. Workshop attendees discussed possible responses to plagiarism. The journal can reject a manuscript and bar future submissions (which can be difficult to enforce); force the author to retract a published paper and place a notice to this effect in the journal; and finally notify the author s institution.

Jean Paul Connerade said that problems have gotten worse due to the ease of cutting and pasting text electronically. He said the publisher s responsibility was to assess scientific accuracy, and not to confirm originality. Regarding citation, students are no longer instructed in correct practices and it is sometimes hard to know what rules an author is following. Standards vary by cultures and generations. In Eastern Europe, professors may commonly think that their students should evermore reference them when they publish. Even more often it is a lab director who expects to be referenced or automatically included as an author. Connerade pointed out the resources required in a system of investigation and denunciation. He stated his opposition to placing responsibility for dealing with plagiarism on the publisher.

Liu Jixing described a case that arose in China some 15 year ago, wherein a Chinese author plagiarized and translated out of a Swiss journal. When this came to light, the author s other papers were examined and many were found to be plagiarized. It can be particularly difficult to find translated plagiarism between journals of lower visibility. The author was fairly highly placed, Liu said, so pursuit of the resulting case and imposition of punishment required determination on the part of the community.

Rohini Godbole described sending a paper to a referee who returned it, saying that not only was it plagiarized from of his own work and but that he had already been asked to review the paper by another journal! More journal to journal communication and cooperation would be helpful, she said. The author was contacted but did not respond, although he did submit again. He was contacted again at that time and has not been heard from since. She described several other cases involving authorship questions and serial publication and how her journal responded.

In conclusion, preservation of confidentiality during investigation, cooperation between journals, difficulty of adjudication, and development of common policies were all thought to be important points.

Manuel Cardona said that journals could require a statement from each author on a paper to the effect that he or she had read the paper, or that the corresponding author provide these. Weinberg pointed out that his journal regularly publishes papers with 500 authors. Requiring such statements would be onerous and still might not solve the problem.

And how big is the problem anyway? Carl Bender said that of some1000 papers he had refereed only two were known to have problems. Does that low rate truly justify a major response? Connerade said that the process of detection is highly random and there are likely many cases that go undetected. Godbole said that Paul Ginsparg had a student working on duplication detecting software, which wouldn t catch stolen ideas but would catch more plagiarized text. Cerdiera added that the funding agencies should take more responsibility and that the pressure to publish is part of the problem.

Peter Knight said it is one matter to take action on a case, but what about future submissions from the same author? Is he or she banned? The legal implications are considerable, especially if any information were to be shared between journals. Regarding the possibility of being sued, Mukhi said that the website regarding the Rajput/Joshi case had been set up in such a way that those who viewed it could draw their own conclusions and then sign on if they chose. It has been very effective.

The session on **duplicate submission/publication** was chaired by Diane Sullenberger. She said that the role of the journal editor is to ensure high quality content and that this includes setting and maintaining ethical standards. There are many varieties of duplicate submission and publication. Some overlap between publications is acceptable provided that it is not excessive and that the author discloses it and provides cross referencing. Practices vary between scientific fields and from journal to journal. She proposed a three point test for originality when a version of a submitted article has previously appeared. If, in this first appearance the article gave

sufficient detail to allow replication, was peer reviewed and also publicly accessible, the second submission cannot be considered original. Duplicate publication wastes resources, inflates the literature, damages journals, and violates copyright laws. In some countries however, duplicate submission is driven by the fact that salary is based directly on the number of papers published.

Over eight years in Sullenberger s journal, the Proceedings of the National Academy of Sciences (PNAS) there have been 8 cases, 2 involving overlap with a symposium paper, 4 of duplicate submission and 2 of duplicate publication. Inter-journal cooperation is essential to prepare appropriate retractions and clarify the scientific record, even though journals don t always agree on the severity of the crime. In PubMed s18 years and over 10 million articles, there have been 454 cases of duplicate publication. The rate is low, roughly 1.4%, but increasing.

What steps to take? Some options are to confront the authors and ask for an explanation, send a letter of reprimand and warning, retract the paper and publish a notice, ban the author from future submissions, contact the author s institution, and report the case in the journal or to an ethics board. Hildebrandt pointed out that duplicate submission is nearly invisible and the number of undetected cases is probably high. Journals could collect and share the names of authors who are known to have engaged in this and other unethical behavior. Questions of how to punish and sanction authors are at the crux of the issue. Visible and unpleasant consequences will help to deter further misconduct and to instruct the community. The problems are not discrete either: duplicate submission turns into serial publication, and plagiarism merges into inadequate referencing or questionable authorship practices. The public, on the other hand, is likely to confuse scientific error with fraud. It s not unethical to be wrong as long as errors are admitted and corrected quickly when they are known.

Peter Knight chaired the session on **authorship**. The ethical issues generally fall on one of two sides: improper inclusion of an author (often a department or institution head) and improper exclusion of an author, commonly a junior researcher. These practices are sometimes a matter of institutional policy and sometimes a cultural phenomena and are difficult for individuals to challenge. In the most egregious cases the senior person included as an author is unaware of the paper and/or isn t qualified in the field it treats. Occasionally the other author/s want to include the senior person, if it is thought that the name will improve the chances of the paper being accepted. In other cases a famous scientists name has been included as author without his/her knowledge or consent. On the other side, a junior researcher who clearly deserves authorship is omitted and may be afraid or unable to insist on being included. Some departments expect a certain number of solo-authored papers as criteria for promotion.

Students (and some faculty) need to be instructed as to what an author is and how to define authorship fairly. Generally, an author is someone who has contributed significantly to the research being published. If the definition is very much more specific it may become counter productive by seeming to include or exclude authors inappropriately.

It might be possible to notify all authors when a submission is received and provide access to it. If the roles of the co-authors were included in the paper, gift authorship would be less common. Weinberg said that the authorship concept in experimental high energy physics is unclear and the

amount of work and text required to verify and state what each of the many authors did would be prohibitive. Ziemelis said that *Nature* now allows a summary of authors contributions but not many are included.

John Enderby addressed the group on the **responsibility of institutions.** Some would argue that scientific papers are a mythical reconstruction of what actually happened so perhaps they are always a misrepresentation of scientific practice. Some say that fraud in science doesn t matter because it will be uncovered eventually and only the perpetrator will suffer. Others point out that fraud erodes public support for science and that colleagues, institutions and journals are harmed by it, not to mention the waste of effort and resources. Responsibility is widely spread among institutions and organizations but fundamentally it is up to individuals to maintain a high ethical standard.

He touched upon the several types and degrees of misconduct, from outright fraud to dubious practices such as fudging or otherwise manipulating data, reporting only good results, and exaggerating claims. There are also a range of honest mistakes, including unrecognized systematic errors, ignorance of the literature and misunderstanding of basic physics. And what can be done with anomalous data points, anyway? If the experiment is repeated, new ones may arise. A reason may be found to reject them or a means to process them away, or they can be reported without explanation, thereby raising doubt in the mind of a referee.

How widespread is misconduct? Half of faculty at institutions in the UK and the US report being exposed to some measure of misconduct. The behavior is driven by vanity, peer pressure, financial gain, and misplaced loyalty concerns that aren t likely to go away anytime soon. In biomedicine, human subjects, commercial pressures and difficulty repeating large scale studies are additional complicating factors. Fraud is a symptom of an underlying problem, and tinkering with policies won t affect it. What is needed is a fearless and dispassionate investigation into problems in the practice of science.

Enderby referred to the Nolan report, a review conducted in the UK in recent years on standards of behavior in public life. It identified the following individual and institutional principles for public life: selflessness, integrity, objectivity, accountability, openness, honesty, and leadership. http://www.archive.official-documents.co.uk/document/parlment/nolan/nolan.htm. Leadership is what defines institutional culture. It ensures that explicit policy and systems are in place to guide good practices, and that procedures and proper records are maintained for adjudication purposes if a problem should arise. It protects whistle blowers as well as the rights of the accused, and seeks objective and independent advice when necessary. Along these lines, the IOP has set up an ethics committee with some external membership, set clear guidelines for the conduct of physicists, and invited the accreditation committee to require departments to have a curriculum of instruction in ethical behavior.

Guest authorship is one problem that can be addressed at the workshop, it was thought. When an author s name is on a paper, he or she receives credit and also must take responsibility or blame if anything is amiss with it. Putting one s name on a paper is akin to signing a check: the author name attests to the value and correctness of the work. A clear and internationally accepted

definition of an author would help, and acceptance of certain responsibilities could be integrated into the copyright form that many authors sign before their works are published.

Lee Schroeder chaired the session on fabrication of data. He began this important session with a recounting of the incident in his department regarding element 118. Three events in a run conducted in 1999 gave evidence of 118, and the resulting article was published in Physical Review Letters. Two external attempts to replicate didn't use identical experimental conditions, and another at LBL was inconclusive. The red flag didn't go up just yet, however, since so few events had been observed originally. At this point a review was conducted of the steps taken in the original run, and in a final attempt at confirmation, a single 118 event came and went. The 1999 data was then reexamined and the original events were no longer in evidence. It was clear that something was wrong.

Schroeder commissioned an internal technical group to look at the data and granted them access and authority to investigate. Simultaneously, the process of retracting the PRL paper got underway. The technical group first verified the accuracy of the software that produced the 1999 data. Then, using sophisticated computer techniques, the data manipulation came clearly and convincingly to light as did the fact that only one person had been responsible for the key data. The committee made its report to Schroeder, and at this point the University policies and procedures were activated. Accordingly, Schroeder asked an externally chaired and internally staffed committee to look at the allegation of misconduct and it determined that there were indeed grounds on which to proceed. Eventually the PRL paper was withdrawn by "all but one of the authors."

As a primary lesson to be drawn from the incident, Schroeder emphasized that extraordinary claims need extraordinary proof. Defining extraordinary may not be so simple, however. He also urged a critical examination of the conduct of operations and scrutiny (if not elimination) of vulnerable points in the procress of research, such as passage through a single individual. He emphasized the need for early, formal education in scientific integrity.

Peter Adams discussed the Schoen case and its associated papers. The Beasley Committee only examined the papers in *Physical Review Letters* that were brought to their attention; there were at least four papers in *Physical Review B* that were not looked at although they were clearly in the same category. Adams said he had been in contact with the authors of these remaining papers and all but Schoen agreed to have them withdrawn; on one paper Schoen was the sole author and that paper remains posted. The citation trail on all papers has also been terminated. Schoen s co-authors may have been asleep at the wheel but on the other hand the natural tendency is certainly to trust one s colleagues.

Julia Fahrenkamp-Uppenbrink remarked that the article length restriction inhibits replication for some *Science* articles, but that it is now possible to post backup data online. It isn t clear that journal editors or peer reviewers can innoculate the community from fraud or even reliably detect it. The peer review system was never designed or intended to perform this function. Everyone in the community was dazzled and taken in by the results reported in the two recent cases of data fabrication. Karl Ziemelis said that *Nature* staff examined the original handling of the tainted

papers and there was nothing untoward at the time. The papers were processed and evaluated in a manner indistinguishable from other normal papers. Referees were unanimous in recommending publication. Some questioned the interpretation of data but none doubted its quality. A more vigilant and critical approach on the part of journals and referees is nonetheless likely in the future. The process of retraction is better documented as well. What could also be improved is the vigor of the response once an investigation is underway. Adams agreed that *Physical Review* did not detect anything irregular in the peer review process for these papers, although the prestigious authors and institutions lent a credence. When the real situation became clear, he said, it was good that the several journals involved shared their plans and information and were able to present a strong and united front.

Cardona said that when questioned about a single curve graph, Schoen had substituted another with three curves such that the noise was hidden. *Science* should have noticed this, he said, but it was a failure of the whole community. Blume said that the investigation at Lucent was carried out in accordance with federal policy even though the corporation wasn t bound by it. Schroeder said that sharing information and building policy was the way to insure that these things don t happen again. Adams said he didn t doubt that a similar case could occur again and in fact that he believed fraud on a smaller scale is surely going on.

In response to an assertion that an external investigation committee should have been recruited in the LBL case, Schroeder described again the three steps that he took. First the group itself tried to replicate but could not, then the original data was examined and nothing came to light, and finally an internal technical committee was assembled and their investigation showed that the data had been manipulated. An external group might not have had the technical expertise required, although he recognized that in some cases an outside committee is needed. A single individual was the source in each of the two cases under discussion, and the structure wherein the work passed through those single points allowed the fraud to be perpetrated. When a team is involved at each step, there is less risk.

Liu observed that paying more attention to experimental data is the common lesson to be drawn. Supervisors should rigorously review and verify the results of students and junior researchers. Fahrenkamp-Uppenbrink pointed out that the rush and pressure to publish makes it less likely that data will be critically examined. It would help if there were standards of reporting from authors, for example on how images were manipulated. Ziemelis said that *Nature* did reflect on the its review processes but without finding any fundamental flaws.

Pieter Bolman chaired the next session, on the **responsibilities of journals to one another**, and the conversation followed closely that of the previous session. Jerome said that more authors should have shared the blame with Schoen. His results weren t surprising, as they were ostensibly based on well established physics. Lucent was happily and readily accepting of the work, as were the journals to which the papers were sent. Another factor must be the enormous pressure on young researchers to publish well and frequently at all costs.

Before *Science* and *Nature* will send a submission out for review, the editors make an initial judgement about whether the paper would be good for the journal were it to be accepted.

Rejection without review is the norm. On the other hand the editors are not bound by a referees positive report either. Editors of other journals generally get a paper refereed first before a judgement is formed about whether the paper would be good for the journal. Overall, the number of submissions to journals is huge, with impact factor imperialism a driving force. A budget of one paper per person per year was lightly suggested.

Minoru Takahashi described an unusual case of two authors who were separated and unable to find and contact each other and who each submitted the paper simultaneously, one in China and the other in Japan. Bolman said that the randomness of uncovering misconduct could be reduced if journals had a neutral way to share information. Perhaps the Ginsparg archive could take this role. When plagiarism or data fabrication is discovered in a published paper, it is often best to allow the paper to remain posted, albeit with a prominent notice affixed, so that the history is clear. Such papers can be excluded from citation data. Bolman asserted the existence of editorial misconduct, the penalty for which is simple dismissal. Furthermore, the community tends to rally around the editor in spite of what he/she might have done.

Donald Levy questioned the 1% figure that had been mentioned in the course of the workshop as representing the level of misconduct in scientific publishing. Jerome said he had seen this number posited in Nature and Science, and that he thought the percentage was probably even lower.

Cardona said that Schoen had done his doctorate at the University of Contanz and that he published unusually heavily there too. An investigating committee found that data had been manipulated in some of those papers, but nothing that couldn t be corrected by means of an errata. Connerade asked whether peer reviewed had failed in these cases. The pressure on individuals and institutions to publish has been covered already, he said, but refereeing hasn t been discussed. If it were given more value and used in assessment for advancement, referees might be able and willing to invest more of their time and do an even better job than at present. Blume said that referees sometimes request a letter to their department chairs, and that raising the profile of referees was a good idea. Physical Review is developing a program to send letters of recognition to the busiest and best referees. Thom Von Foerster said that one of the AIP journals publishes an annual list of those who refereed. Godbole called for cross-journal accessibility of publishing records for editors, but Liu wondered about this being an invasion of privacy if not a violation of the law. Mukhi proposed that each journal link any published retractions to a central site.

14 October

Richard Smith, Editor of the British Medical Journal, opened the second day of the workshop. He gave a lively account of the better known cases of **misconduct in biomedical research** in recent years. Failure to take action when fraud is noted is punishable, he said, to counteract a tendency in medical research to turn a blind eye to misconduct. One reason for this may be the difficulty in investigating fraud in biomedicine. The raw data may indeed be provided but the amount of material and the time needed to analyze it can be staggering, so any excuse will do to avoid this. He remarked that peer review in medicine is open; identities of referee and author are known to

each party.

A definition of misconduct would help but may not be achievable. Smith estimated that about a fifth of all biomed papers contain some degree of redundant publication, and about a fifth of all authors have done little or nothing towards the paper for which they are receiving authorship credit. In a few such cases recently, the junior author s work turned out to be fraudulent and the guest author s career suffered as a result.

In the UK, a Committee on Publications Ethics, (appropriately, COPE), has been established as a resource for biomed editors struggling with ethical issues. COPE s deliberations include a collection of best publication practices and reports of anonymised cases that the Committee considered. http://www.publicationethics.org.uk/ A number of the government agencies that fund biomedical research now require that recipient institutions have a policy and method of dealing with misconduct. An international ethics committee exists within UNESCO, and might be able to set ethical standards across all the sciences and humanities. Such a process would be slow to come about but could have a broad and enduring effect.

Thom Von Foerster moderated the next session, a panel discussion of **conflict of interest**. Fortunately, financial conflicts of interest in physics research are not as common as they are in other fields, but they do arise in applied physics and in instrumentation when principles of corporations write papers about their products without disclosing those connections. Conflicts of interest extend well beyond matters of financial gain, although these are the most commonly cited. Some conflicts can be quite subtle and a researcher may be unaware that they exist. A researcher may have an unconscious and perhaps unscientific prejudice about what methods or approaches are best, or may be on one side of two disagreeing factions or positions. Religious and political bias can enter into research and evaluation decisions as well.

Disclosure of any conflicts is thought to be a panacea, and indeed very often it is all that is necessary. Problems arise from non-disclosure but also from inconsistent practices in disclosure. Sometimes a declaration is voluntary and not formally required. Authors may be confident that they are not affected or conflicted and so choose not to make a statement, but this practice tends to cast doubt when conflict is disclosed. At present, a disclosure of conflict of interest may have the effect of biasing the reader regarding the objectivity of the report exactly the opposite effect from that intended. When there is a conflict of interest or even the possible *appearance* of one, it should be revealed. A consistent practice and perhaps a stricter policy is necessary, although authors are already heavily burdened with requirements.

The next session, on **referee misconduct**, was chaired by Katsu Ushioda. Some of the issues in this area are closely related to those in conflict of interest. Referees occasionally recuse themselves if they feel that a factor related to the author or research prohibits them from making an objective report. Referee misconduct is one of the most difficult types of misconduct to detect, because of the need to preserve confidentiality. There are active forms of referee misconduct, such as issuing a negative report and then using the privileged information contained in that paper for his or her own purposes. More passive and nebulous forms of misconduct on the part of a referee are excessive or deliberate delay in reporting, vague reports or faint praise, and abusive

language in reports. All the journals represented at the workshop are dependent on the good will of referees. A single high profile case of a referee being punished for misconduct could have a significant chilling effect on the willingness of reviewers to provide their frank and honest opinions. The constraints of confidentiality do dictate however that this kind of misconduct is best investigated and adjudicated by journal staff.

Refereeing is generally thought to be a thankless job, and changing that would encourage good behavior. Referees should be thanked often, formally and informally. Records on their activity are useful and can be referred to in any expressions of gratitude. Increasing the size of the referee pool helps to relieve the burden and allow reviewers to report more thoughtfully.

Hilda Cerdiera lead the session on **different international views of misconduct**. It had been made clear in the course of the workshop that there was little difference of opinion from country to country as to what constitutes misconduct. Standards are high worldwide. Cerdiera said that authors from developing countries are concerned at times with the actions of editors and referees. Editors appear to want a more enthusiastically positive referee report to accept these papers than they would if the work came from a prestigious western institution. Referees seem often to press for more references to themselves in the papers they review for developing country authors, who are not in a position to refuse. Direct evidence for these impressions is hard to come by, but the widespread impression persists that papers from developing countries have to meet a higher standard than those from western countries. Double blind refereeing was suggested as a possible remedy.

At times referees seem unwilling to overlook substandard English and judge instead the quality of the physics. There s a degree beyond which the physics is incomprehensible due to the poor English, but referees should be willing to make an effort. It is hard for international authors to overcome the impression of bias. A strong international representation on editorial boards and in journal editorial staff helps to dispel this.

The final session, chaired by Jim Tsang, concerned **punishment of offenders**. There are a variety of steps that can be taken by the parties involved: colleagues, employers, journals, agencies and professional societies. In some countries and/or fields, there is an obligation to report ethical violations and failure to do so becomes misconduct itself. Journals can post retractions and impose an electronic pillory on authors of plagiarizing or otherwise tainted papers. They can also refuse future submissions, but an air tight system to enforce this step is elusive for some publishers. Professional societies may revoke membership, provided that there are grounds for this step, i.e. that the member agreed to uphold certain ethical standards as a condition of membership. Funding agencies may demand a return of the grant money that paid for the fraudulent research and may well deny future proposals.

In the most egregious cases, institutions may dismiss employees known to have engaged in serious misconduct. Some institutions may be reluctant to take action against their employees. The risk of litigation is a factor as is the concern that it will reflect poorly on the institution itself. Investigation can be very problematical, particularly when dealing with large amounts of raw data. But for a scientist the loss of reputation is the very worst punishment, even worse than loss

of employment.

In conclusion, the rapporteurs presented their reports and the workshop adjourned at approximately 4:00.