

БЛЕСК И НИЩЕТА ФОРМАЛЬНЫХ КРИТЕРИЕВ НАУЧНОЙ ЭКСПЕРТИЗЫ¹

Фрадков А. Л.²

*(ФГБУН Институт проблем машиноведения РАН,
Санкт-Петербург)*

В статье обсуждаются проблемы сочетания формальных и неформальных критериев научной экспертизы на примере конкурса мегагрантов. Описывается попытка выработать сопоставимые критерии для различных областей наук.

Ключевые слова: наукометрия, научная экспертиза, экспертные оценки.

Не многого ль хочу, всему давая цену?

Б. Окуджава

1. Введение

В наш век падения авторитета науки в обществе, когда большинство в обществе, в том числе большинство лиц, принимающих решения, не понимает большинства результатов деятельности ученых, возникает естественный вопрос: а зачем нужны эти ученые, зачем нужна наука вообще?

Осмелюсь предположить, что из немногих вариантов ответа особого внимания заслуживает следующий: наука – это способ сохранения в обществе понятий об истине и лжи, а ученые – это люди, которые способны отличить истину от лжи в окружающей

¹ Сокращенная версия заметки опубликована в газете «Троицкий вариант», 29.01.2013, №122.

² Александр Львович Фрадков, доктор технических наук, профессор (fradkov@mail.ru).

нас реальности. Сохранение такой группы людей в обществе, в котором ключевые слова «прогресс» и «мораль» вытеснены словом «интересы», важно, чтобы общество не потеряло стратегические и моральные ориентиры.

Но чтобы подобные уникальные качества сообщества ученых не обесценивались, понимание научной истины и морали внутри группы должно быть безупречным и очевидным для всех – иначе ученые потеряют доверие общества. Вот почему так важны понятность и безупречность процедур экспертизы – механизмов установления научной истины. Безукоризненность экспертизы важна и при рецензировании научных публикаций, и при отборе научных и технологических проектов и, наконец, при защите диссертаций – оценке квалификации научных работников.

Во всех перечисленных случаях задача упрощается, если рассматриваемые объекты экспертизы принадлежат к одной области наук или технологий, т.е. критерии их оценки могут быть относительно однородны. Но что делать, если стоит задача отобрать для премирования лучшую из десяти работ по математике, химии, экономике, медицине и истории? А если работ не десять, а сто или тысяча?

В заметке обсуждаются проблемы критериев экспертизы на примере недавно завершеного конкурса на продление первой волны мегагрантов, где предполагалось продлить примерно половину из поданных заявок. Относительным упрощением задачи в этом случае являлось то, что на неформальном уровне имелся консенсус относительно критерия успешности проекта. Именно, требовалось проверить достижение основной цели: создать за два года устойчиво работающую лабораторию мирового уровня.

К сожалению, в настоящее время российская наука резко поляризована, разделена на международную и региональную (местную, провинциальную). Казалось бы, все признают важность интеграции российской науки в мировую, признают, что «наука не знает ни рас, ни границ, ни конфессий, ни пола, ни возраста, ни национальности» [4], РФФИ уже недвусмысленно требует проводить

исследования, соответствующие мировому уровню и т.д. Однако все еще находятся «патриоты», упрекающие чуть ли не в измене Родине тех, кто стремится оценивать достижения российской науки, сравнивая ее с мировой, и поддерживать публикации в престижных международных журналах [2, 9].

В данной работе мы избегаем обсуждения указанной проблемы, не вдаваясь в политику и ограничиваясь лишь анализом вопросов экспертизы для конкурса мегагрантов. Будем рассматривать конкретный вопрос: как, по каким критериям определить, достигла ли лаборатория мирового уровня? Совет по грантам при Правительстве РФ должен был выработать эти критерии и принять решение на их основе.

2. Критерии мирового уровня лаборатории

Неудивительно, что речь о критериях мирового уровня зашла на первом же заседании Совета. Удивительно то, что у собравшихся специалистов не было серьезных разногласий по ним. Это важно потому, что предстояло оценить в сопоставимых терминах 37 поданных заявок на продление из 17 различных областей наук. Чисто формальные критерии, увы, не работают, поскольку формальные показатели журналов, традиции публикационной активности и цитирования различаются в разных науках. Кроме того, уровень лаборатории зависит от уровня ее инфраструктуры, который трудно оценить по формальным критериям. Поэтому первоначальную оценку проектов выполняли эксперты.

Как соединить достоинства формальных показателей и экспертных оценок? Казалось бы, просто: каждую заявку оценивает множество независимых экспертов (в конкурсе мегагрантов требуется участие двух российских и двух иностранных экспертов на каждый проект). Каждый эксперт ставит балл от 1 до 5 по каждому показателю, а затем баллы экспертов суммируются с весами, задающими относительную важность показателей (см.

далее п.3).¹ Совету остается только упорядочить заявки по убыванию суммы баллов и выбрать для продления первые N заявок, где N определяется либо заданным числом победителей, либо заданным бюджетом конкурса. Однако формальный подход имеет целый ряд недостатков. Вот основные из них:

1. Никакой эксперт не может быть абсолютно объективным, т.е. его оценка имеет погрешность и отличается от истинной (предполагаем, что таковая существует). Значит, упорядочивание по баллам также отличается от истинного. Поэтому правильное отобрать заявки «с запасом», т.е. имеющие не только «проходной» балл, но и «полупроходной». «Полупроходные» заявки следует рассматривать отдельно, анализируя оценки экспертов. При этом важно, чтобы полупроходная зона не содержала слишком много заявок, так как иначе не хватит времени на их детальное рассмотрение.

2. Еще более опасными могут быть погрешности, порожденные необъективностью отдельных экспертов. Мы столкнулись с несколькими случаями, когда трое из четверых экспертов дают примерно одинаковые суммы баллов, а оценка четвертого резко отличается. Причиной может быть конфликт интересов (четвертому выгодно, или наоборот, невыгодно ставить высокую оценку). Или наоборот: четвертый эксперт нашел ошибку или слабость заявки, которую не заметили остальные.

В обоих случаях требуется детальный анализ заявки одним или несколькими членами Совета, желательно, до решающего заседания. Именно так все и происходило. Члены Совета, заметив неоднозначность оценки, вникали в заключения всех экспертов, а иногда смотрели саму заявку, причем не только из «своей» области знаний. Дополнительную информацию давало мнение руководителя группы экспертов, играющего роль моде-

¹ Хотя на первый взгляд суммирование оценок кажется упрощенным способом учета мнений нескольких экспертов, известны исследования, обосновывающие именно такой подход [11, 13]. Он широко распространен в экспертизе проектов и применяется многими фондами, включая РФФИ.

ратора и составляющего общее заключение по рецензиям четырех экспертов для всех проектов из данной области наук. Такое заключение может помочь Совету, но может и внести дополнительную погрешность. Во всех спорных случаях члены Совета обращались к первичным показателям заявки, прежде всего – к публикациям. Кроме уровня публикаций оценивалась место их выполнения: если большая часть работ была выполнена за рубежом, то мирового уровня достигла не российская лаборатория, а зарубежная. Наконец, важную роль играла развитость инфраструктуры созданной лаборатории. Перечисленные неформальные рассмотрения дополняли формальный показатель «сумма баллов экспертов» и позволяли снизить влияние недостатков, присущих формальной экспертизе. Для продления были отобраны 24 проекта, причем, чтобы уложиться в допустимый объем финансирования, сумму, выделяемую на каждый грант, уменьшили на 10%.

3. О робастности ответа относительно весов

В ходе работы произошел следующий казус, поучительный для дискуссии о соотношении формальных показателей и экспертных оценок. Перед итоговым заседанием Совета аналитическая группа МОН подготовила ведомость оценок и заключений экспертов по каждой заявке на продление. Оценки выставлялись по десяти частным критериям, разбитым на две группы. Первая группа оценивала значимость достигнутых результатов (шесть критериев, в том числе соответствие результатов плану, уровень публикаций, уровень выступлений на конференциях, кадровый состав лаборатории), а вторая группа – перспективность дальнейших исследований по проекту (четыре критерия, в том числе актуальность, достижимость целей и т.п.). Каждому частному критерию соответствовал вес, с которым значение критерия входит в итоговую оценку. В первой группе у первых двух критериев был вес по 30%, а у оставшихся – по 10%, а во второй группе веса всех частных критериев были по 25%. Далее, вклад первой группы умножался на 0,7, а вклад второй группы

– на 0,3. На предыдущем заседании Совета веса обсуждались, менялись, а в одном случае даже голосовались. Так вот, в первом варианте подготовленной ведомости оценок, разосланном членам Совета накануне, по недосмотру все веса были взяты равными, и именно этот вариант использовался нами для анализа. Однако потом ошибка была исправлена и за 2-3 часа до заседания была разослана исправленная ведомость, которая раздавалась и на самом заседании. Так вот, курьез заключается в том, что изменение весов практически не повлияло на порядок расположения заявок в списке по убыванию рейтинга! Сами итоговые суммы, конечно, изменились, а вот порядок остался практически тем же, за исключением двух-трех «полупроходных» заявок, которые все равно бы рассматривались на заседании. На мой взгляд, это объясняется тем, что для рассмотрения на заседании отбирались заявки с достаточно высокими баллами по всем показателям, т.е. приходилось отбирать из хороших «очень хорошие» проекты, вследствие чего показатели оказывались коррелированными и умеренные изменения весов слабо влияли на результаты отбора.

Отметим, что в существующей теории экспертных оценок [5, 6] показывается, что робастность выводов повышается, если критериев достаточно много, причем коррелированность критериев не играет существенной роли. На практике, однако, введение большого числа критериев создает проблемы сбора данных, выбора весов и в общем случае не может быть рекомендовано.

4. Сочетание формальных и экспертных оценок

Параллельно с оценкой проектов на продление нам приходилось обсуждать и принимать правила и критерии для нового конкурса (третьей волны) мегагрантов. Здесь возникают те же проблемы, но, поскольку конкурс еще не начался, была возможность заранее принять меры, чтобы правила, по которым эксперты будут выставлять свои оценки (они лежат в диапазоне от 1 до 5), были сопоставимы для различных наук. Для этого было решено по каждому критерию выдать рекомендации, в каких

случаях эксперт должен ставить высший балл. На первый взгляд, такая задача неразрешима: единые рекомендации для разных наук выдать нельзя. Этот вопрос обсуждался ранее на форуме сайта Scientific.ru, где был предложен описываемый ниже вариант решения на основе нескольких простых формальных критериев.

Конкурсная комиссия конкурса мегагрантов с самого начала конкурсов установила форму анкеты экспертов. Было принято, что критерии оценки заявки делятся на три группы [8]:

1. Уровень ведущего ученого;
2. Уровень (качество) заявки;
3. Уровень подготовленности принимающего коллектива.

Такое деление разумно, поскольку оно выделяет три ключевых, независимых между собой фактора успеха будущего проекта. В первой группе три частных критерия:

1.1. Уровень научных публикаций. По нему оценку «5» предлагается ставить при обязательном выполнении двух условий:

- 1) ведущий ученый за последние 3 года ежегодно публикует не менее 1-2 статей в ведущих международных журналах;
- 2) ведущий ученый имеет индекс цитируемости или индекс Хирша не менее 40% от показателей лидеров по данной области науки.

Заметим, что определение того, является ли журнал ведущим и какова цитируемость лидеров в данной области наук, полностью отдается на откуп экспертов. Для ориентировки экспертов мы предложили ориентироваться на список ведущих журналов, принятый при расчетах стимулирующих выплат в МГУ и содержащий верхние 25% журналов по импакт-фактору из числа индексируемых Web of Science, а также все журналы из базы Thomson Reuters Arts & Humanities Citation Index [8].

Следует заметить, однако, что список МГУ подходит далеко не для всех областей наук. Его не признают математики, некоторые информатики и многие гуманитарии. У математиков зачастую считаются престижными другие журналы, не отличающиеся высокой цитируемостью [2]. У компьютерщиков мериллом успеха являются не журнальные статьи, а приглашен-

ные доклады на нескольких престижных конференциях. В классической филологии же и смежных областях ценятся не журнальные статьи, а монографии и, уж конечно, не на английском языке.

1.2. Опыт ведущего ученого по руководству научным коллективом. Здесь оценка «5» ставится, если выполнены хотя бы 3 из следующих 4 требований:

1) ведущий ученый создал хотя бы один международный научный коллектив и управлял его деятельностью в течение хотя бы 5 лет;

2) ведущий ученый в течение не менее чем 5 лет руководил хотя бы одним международным грантом или проектом;

3) ведущий ученый управлял деятельностью хотя бы одного национального научного коллектива в течение хотя бы 5 лет;

4) ведущий ученый руководил хотя бы двумя крупными национальными грантами или проектами в течение не менее чем 5 лет.

1.3. Опыт и возможности ведущего ученого по подготовке научных и педагогических кадров. Оценка «5» ставится, если ведущий ученый за последние 7 лет руководил не менее чем тремя аспирантами (PhD students), два из которых успешно защитились.

Во второй группе четыре частных критерия:

2.1. Актуальность планируемых научных исследований;

2.2. Достижимость заявленных результатов в предложенные сроки и заявляемыми методами;

2.3. Соответствие запрашиваемого финансирования поставленным целям, качество проработки сметы проекта;

2.4. Перспективный облик лаборатории, создаваемой в организации в рамках проекта, через три года.

Здесь нам не удалось дать количественных рекомендаций, и они остались на качественном уровне, так как были предложены министерством [7].

В третьей группе пять критериев:

3.1. Публикационная активность коллектива участников заявляемого проекта;

3.2. Имеющаяся у коллектива участников заявляемого проекта научная инфраструктура;

3.3. Адекватность принимаемых организацией обязательств по созданию лаборатории;

3.4. Кадровый состав организации;

3.5. Роль лаборатории в решении задач организации по ее инновационному развитию.

Удалось квантифицировать только критерии 3.1 и 3.4, а именно:

По 3.1 оценка «5» ставится, если выполнены хотя бы 3 из следующих 4 требований:

1) коллектив за последние 3 года ежегодно публиковал не менее 4-6 статей в международных журналах;

2) коллектив за последние 3 года ежегодно представлял не менее 3-5 докладов на важнейших международных конференциях;

3) коллектив имеет в своем составе участников с индексом цитируемости или индексом Хирша не менее 60% от показателей лидеров по данной области науки в России;

4) коллектив планирует за два года проекта иметь не менее 4-6 статей, опубликованных или принятых к печати в международных журналах, индексируемых в Web of Science или Scopus, в том числе статьи в ведущих международных журналах.

По 3.4. оценка «5» ставится, если выполнены хотя бы 3 из следующих 4 требований:

1) наличие в составе коллектива не менее 20% молодых ученых и аспирантов;

2) наличие в составе коллектива не менее 30% участников, опубликовавших за последние 3 года не менее трех статей в международных журналах, индексируемых в Web of Science или Scopus;

3) наличие в составе коллектива участников, имеющих опыт руководства грантами, государственными и хозяйственными контрактами;

4) наличие в составе коллектива участника, имеющего опыт административного руководства научным коллективом (лабораторией, кафедрой)

Легко видеть, что критерии этой группы требуют, чтобы у принимающего коллектива был определенный уровень, чтобы передовые научные идеи, брошенные в эту почву, смогли взрастить лабораторию мирового уровня за 2-3 года. Конечно, далеко не каждый коллектив готов к решению такой задачи: у существенной части группы должен иметься хотя бы минимальный опыт работы на мировом уровне. Как сказал один из членов Совета по грантам, «нельзя приглашать Гуса Хиддинка тренировать дворовую команду». Это мы и пытались формализовать, хотя бы частично.

После выставления всеми экспертами оценок они будут просуммированы с заранее зафиксированными коэффициентами и дадут формальный показатель качества (балл) заявки. Для принятия решения Совет будет рассматривать, кроме этого числа, еще рекомендацию руководителя группы экспертов по области наук, в сомнительных случаях обращаясь к самой заявке.

5. Заключение-1

Подводя итог, хочется еще раз подчеркнуть, что для объективной экспертизы недостаточно использовать только формальные показатели, так же как недостаточно использовать и только экспертные оценки, причем во всех областях наук. В вечном споре о том, как правильнее решать: по формальным критериям или по экспертным оценкам, нет победителя. Первый опыт работы нового Совета по грантам это еще раз подтвердил. Правильное решение должно основываться на разумной комбинации обоих подходов. При этом важно, чтобы на последнем этапе решение принималось, по возможности, не голосованием, а консенсусом и слабо зависело от выбора коэффициентов на предыдущих этапах. Вопрос в том, удастся ли эти принципы реализовать, когда выбирать придется не из сорока, а из нескольких сот заявок. Ясно, что даже если рассматривать внимательно только «полупроходные» заявки, это потребует серьезных усилий и затрат времени членов Совета.

Хочется надеяться, что Совет справится. А если все-таки нет, и будет ясно, что сделать все области «равноудаленными» не получается? Альтернативой является квотирование по областям наук: число победителей в каждой области должно быть пропорционально числу поданных заявок. А если эксперты находят, что уровень заявок из какой-то области существенно ниже, чем в других? Что ж, и этот перекос можно учесть: надо выделять квоту на область не пропорционально числу заявок, а пропорционально сумме баллов экспертов по этой области. А если потребуется учесть дополнительные факторы: необходимость поддержки регионов за счет столиц или ограничения сверху на число победителей в организациях, где грантов и так много? Думаю, что можно построить систему квот и весов и для этого случая, хотя подобные задачи наверняка нуждаются в дополнительных исследованиях.

6. Заключение-2

За время, пока статья писалась, обсуждалась и рецензировалась, новый конкурс мегагрантов был благополучно проведен. Из 550 заявок было отобрано 42 победителя (первоначально было подано более 700, но часть из них не была допущена к конкурсу по причине формальных нарушений). Описанные выше подходы к экспертизе еще раз подтвердили, на мой взгляд, свою разумность.

Как и ожидалось, при экспертизе большого числа заявок высветились наиболее острые проблемы. Одна из них: разброс мнений экспертов. Когда проводится экспертиза нескольких десятков работ, особых трудностей нет: каждую заявку с большим разбросом мнений экспертов можно проанализировать «вручную» и определить, мнению какого эксперта следует меньше доверять. Однако при экспертизе нескольких сот заявок Совет уже не в состоянии просмотреть все подозрительные работы, и возникает нужда в автоматизированных процедурах. В прошедшей третьей волне конкурса мегагрантов, как и в первых двух, экспертиза проводилась в два тура. Сначала со-

ставлялся «короткий лист» заявок: по общей сумме баллов отбиралось 15-20% заявок для второго тура, а остальные далее не рассматривались. Число заявок во втором туре было около 100, что примерно в два раза превышает предполагаемое число победителей. Такое число, с учетом разных областей компетентности членов Совета, уже реально обработать вручную.

Однако при двухтуровой процедуре у большинства заявок контроль разброса результатов экспертизы не производился, и ряд авторов, получив результаты экспертизы, выразили недовольство тем, что оценки одного из четырех экспертов были несправедливо занижены, вследствие чего они даже не попали в «короткий лист».

С проблемой снижения разброса экспертных оценок (достижения консенсуса экспертов) социологи сталкиваются давно, и существует целый ряд проверенных на практике решений [5, 6, 10, 14]. Метод, используемый в практике экспертизы европейских исследовательских проектов, основан на проведении встреч экспертов, так называемых «консенсусных сессий» [10]. Однако организация таких сессий требует финансовых и временных затрат.

Широкое распространение нашел метод Дельфи и его модификации [6, 14], основанный на многократном обмене экспертами информацией о своих решениях. В рассматриваемом случае наиболее подходящей является модификация, в которой на каждой итерации всем экспертам сообщаются все оценки и их среднее, и предлагается скорректировать свои оценки. При этом экспертам, оценки которых дальше всего отстоят от среднего, предлагается сообщить обоснование своих оценок. На последней итерации для получения окончательной оценки отбрасывают оценку, наиболее сильно отклоняющуюся от среднего (или отбрасывают наибольшую и наименьшую оценки), а остальные усредняют. Известно много математических результатов, дающих условия сходимости итеративных оценок различного рода, если число итераций неограниченно растет [1, 12, 15]. Практически, однако, число проведенных итераций зависит от временных и финансовых возможностей организаторо-

ров экспертизы. В нашем случае, на мой взгляд, существенное улучшение может наступить, даже если выполнить хотя бы одну итерацию метода.

Другой проблемой является различие в количестве и уровне заявок, поданных по различным областям наук. Чтобы обеспечить равномерность конкурса, были определены квоты на число победителей, пропорциональные числу поданных заявок по областям наук. Для учета качества заявок квоты были пересчитаны по итогам рецензирования пропорционально суммам баллов по областям наук. Разница в квотах до и после рецензирования оказалась невелика: была увеличена квота по физике и уменьшена квота по общественным и гуманитарным наукам. Тем не менее, она отражает реальную конкурентоспособность заявок из разных областей, что обосновывает целесообразность коррекции квот.

Заканчивая обсуждение, заметим, что перечисленные в статье проблемы возникают во многих видах экспертизы научных проектов, журналов, оценки эффективности работы институтов, лабораторий и отдельных ученых за заданный период и т.д. Поэтому перечисленные практические приемы принятия решений могут оказаться полезными и в других областях.

Литература

1. АГАЕВ Р.П., ЧЕБОТАРЕВ П.Ю. *Сходимость и устойчивость в задачах согласования характеристик (обзор базовых результатов)* // Управление большими системами. – 2010. – Вып. 30.1. – С. 470–505.
2. ГУСЕЙН-ЗАДЕ С.М. *Повесть об ИСТИНЕ* // Управление большими системами. – 2013. – № 44 – С. 422–435.
3. КОБРИНСКИЙ А.Л. *Мы опять проводим опыты на людях!* – URL: <http://www.smena.ru/news/2013/02/25/21709> (дата обращения: 15.07.2013).
4. КУТАТЕЛАДЗЕ С.С. *Александров и Смирнов, просветители из Петербурга.* – URL: <http://www.math.nsc.ru/>

- LBRT/g2/english/ssk/ad&vi.html (дата обращения: 15.07.2013).
5. ОРЛОВ А.И. *Теория принятия решений*. – М.: Экзамен, 2006.
 6. ПАНКОВА Л.А., ПЕТРОВСКИЙ А.М., ШНЕЙДЕР–МАН М.В. *Организация экспертизы и анализ экспертной информации*. – М.: Наука, 1984.
 7. *Порядок проведения конкурса мегагрантов (3-я волна)*. – URL: <http://www.p220.ru/contest/2012/konkurs.ppt> (дата обращения: 15.07.2013).
 8. *Список Топ-25% журналов по критерию импакт-фактора по версии Thomson Reuters*. – URL: <http://istina.imec.msu.ru/statistics/journals/top/> (дата обращения 15.07.2013).
 9. ФЕДОРОВ Е.А. *Слуги США*. – URL: <http://www.efedorov.ru/node/1644> (дата обращения: 15.07.2013).
 10. ФРАДКОВ А.Л. *Как это делалось в Брюсселе // «Троицкий вариант»*. – 29.01.2013. – №60. – С. 4.
 11. ЧЕБОТАРЕВ П.Ю. *Метод строчных сумм и приводящие к нему модели // Проблемы компьютеризации и статистической обработки данных. Сборник трудов*. – М.: ВНИИСИ, 1989. – №3. – С. 94–110.
 12. Blondel V.D., Hendrickx J.M., Olshevsky A., and Tsitsiklis J.N. *Convergence in multiagent coordination, consensus, and flocking // Proc. of the IEEE Conf. on Decision and Control (CDC-ECC 2005)*, 2005. – P. 2996–3000.
 13. СЕВОТАРЕВ Р.Ю., ШАМИС Е. *Characterizations of scoring methods for preference aggregation // Annals of Operations Research*. – 1998. – Vol. 80. – P. 299–332.
 14. DALEY N., HELMER O. *Experimental Application of the Delphi Method to Use of Experts // Manag. Sci.* –1963. – Vol. 9. – P. 458–467.
 15. DEGROOT M.H. *Reaching a consensus // Journal of the American Statistical Association*. – 1974. – Vol. 69. – P. 118–121.

THE SPLENDORS AND MISERIES OF SCIENTIFIC EXPERTISE FORMAL CRITERIA

Alexander Fradkov, Institute of Problems in Mechanical Engineering of RAS, Saint Petersburg, Dr.Sc., professor (fradkov@mail.ru).

Abstract: Joint usage of formal and informal criteria of scientific expertise for the mega-grants' contest is discussed. An attempt to work out a reasonable approach for evaluation of projects from different scientific fields is described.

Keywords: scientometrics, scientific expertise, peer review.

Поступила в редакцию 03.03.2013.

Опубликована 31.07.2013.