О возможных причинах недооценки рисков гибели человеческой цивилизации

А. В. Турчин

Эта статья посвящена ошибкам в рассуждениях о глобальных рисках. В ней под термином «глобальные риски» имеются в виду возможности катастроф, которые приведут к полному и необратимому уничтожению или вымиранию человечества.

Наши рассуждения о глобальных рисках подвержены тем или иным ошибкам, которые оказывают влияние на конечные выводы этих рассуждений, а, следовательно, и на нашу безопасность. Даже если вклад каждой из нескольких десятков возможных ошибок мал, вместе они могут отклонить вероятностную оценку того или иного сценария в разы и привести к неправильному приложению средств обеспечения безопасности. Специалист в области глобальных рисков должен быть в курсе этих подводных камней. В этой статье предпринимается попытка составить список таких ошибок. Использованы работы зарубежных и российских исследователей, а также авторские наработки. Базовым текстом по проблеме является статья Элизера Юдковского «Систематические ошибки в рассуждениях, потенциально влияющие на оценку глобальных рисков», которая выходит в Оксфорде в 2007 г. в сборнике «Риски глобальной катастрофы» и в соответствии с любезным разрешением автора была переведена мной на русский язык. Данный список не заменяет эту статью, в который приведён математический и психологический анализ ряда приведённых здесь ошибок. Однако многие описания ошибок взяты из другой литературы или обнаружены самим автором. Анализ возможных ошибок в рассуждениях о глобальных рисках является шагом на пути к созданию методологии работы с глобальными рисками, а значит, и к их предотвращению.

Цель работы — свести возможные ошибки в удобный и структурированный список. При этом максимальное внимание уделено полноте списка, а не доказательству каждого отдельного пункта.

Данный список не претендует ни на полноту, ни на точность классификации, и некоторые его пункты могут оказаться тождественны другим, но выраженным иными словами. Подробное разъяснение каждого отдельной возможной ошибки в оценке риска заняло бы весь объём статьи. (См. например, статью «Природные катастрофы и антропный принцип» в этом Сборнике, где одна из приведённых в настоящей статье возможных ошибок разбирается довольно подробно).

Вместе с тем важно помнить, что ошибкам в рассуждениях в той же мере свойственная патологическая самоорганизация, как и ошибкам и цепочкам событий, которые приводят к реальным катастрофам. Это означает, что даже небольшие ошибки, приводящие к небольшому отклонению оценок, имеют тенденцию зацепляться одна за другую, взаимоусиливаясь, особенно при возникновении положительной обратной связи с ними.

Ложный вывод часто становится интеллектуальным источником катастрофы. Нетрудно проследить на примере реальных аварий, как ошибочные рассуждения пилотов самолётов приводили к катастрофам, и даже обозначить, какие именно ошибки в рассуждениях они совершили.

Можно сказать, что почти любая катастрофа происходит из-за человеческих ошибок. Эти ошибки хронологически выстраиваются так: вслед за ошибками в рассуждениях о возможностях идут ошибки в конструировании, в «предполётной» подготовке», в пилотировании, в управлении критической ситуацией, в устранении последствий аварии и в анализе её причин. Наши рассуждения о глобальных рисках в основном относятся к первой стадии, к рассуждениям о возможности и предварительной оценке вероятностей тех или иных рисков.

Нет смысла выстраивать стратегию противостояния глобальным рискам до того, как определились приоритеты. Соответственно, описываемые в данной статье ошибки относятся, в первую очередь, к самой ранней фазе противодействия глобальным рискам. Однако они могут проявлять себя и позже, на стадии конструирования механизмов зашиты и принятия конкретных решений. Тем не менее, в этой статье не ставится задачи развернутого анализа ошибок возможных на более поздних стадиях защиты от глобальной катастрофы.

Есть также вероятность, что некоторые описания ошибок, которые приводятся в этой статье, могут оказаться объектами неверного понимания автора — т. е. тоже быть ошибочными.

И нет ни малейших сомнений, что этот список не полон. Поэтому данный список следует использовать скорее как стартовую площадку для критического анализа любых рассуждений о глобальных рисках, но не как инструмент для постановки окончательного диагноза.

При этом нельзя забывать, что навык обнаружения ошибок в чужих рассуждениях может сделать субъекта менее восприимчивым к новой информации. Также опасно избирательное применение этого навыка, т. е. применение его в большей мере к теориям оппонента, чем к собственным.

Ошибки разделены на следующие группы:

1. Ошибочные представления о роли ошибок.
2. Ошибки, возможные только при анализе и оценке глобальных рисков в силу их специфики.
3. Ошибки, возможные относительно оценки любых рисков, применительно к глобальным рискам.
4. Общие логические ошибки, могущие проявиться в рассуждениях о глобальных рисках.
5. Специфические ошибки, возникающие в дискуссиях об опасности неконтролируемого развития искусственного интеллекта (ИИ) (а также специфические ошибки в рассуждениях о нано-, био- и других прорывных и опасных технологиях). Эти ошибки в данной статье не рассматриваются.

Ошибочные представления о роли ошибок

Опасная иллюзия состоит в том, что ошибки в рассуждениях о глобальных рисках или невелики, или легко обнаружимы и устранимы. Корни этой иллюзии в следующем рассуждении: «Раз самолёты летают, несмотря на все возможные ошибки, и вообще жизнь на Земле продолжается, то значение этих ошибок невелико». Это аналогия неверна. Самолёты летают потому, что в ходе их эволюции, конструирования и испытаний разбились тысячи машин. И за каждой этой аварией стояли чьи-то ошибки, которые каждый раз учитывались и, в целом, не повторялись. У нас нет тысячи планет, которые мы можем разбить, чтобы понять, как нам правильно обращаться с взрывоопасной комбинацией био-, нано-, ядерных и ИИ технологий. Мы не можем использовать и тот факт, что Земля

ещё цела для каких-либо выводов о будущем (см. статью «Природные катастрофы и антропный принцип» в этом Сборнике), потому что нельзя делать статистических выводов по одному случаю. И, конечно, особенно потому, что будущие технологии могут принципиально изменить жизнь на Земле. Итак, мы лишены привычного способа устранения ошибок — проверки. И, тем не менее, именно сейчас нам важнее всего, как никогда ранее в истории человечества, не ошибиться.

Незавершённость этого списка ошибок. Возможно, что есть ряд ошибок, которые проявляются только в рассуждениях о глобальных рисках, и которые пока не обнаружены, но полностью меняют весь ход рассуждений.

Точно также нельзя утверждать, что все исследователи допускают все эти ошибки. Наоборот, большинство этих ошибок, вероятно, самоочевидны большинству исследователей — или вообще не кажутся ошибками. Однако есть шанс, что какие-то ошибки пропущены.

Под «ошибками» в контексте данной статьи рассматриваются не только нарушения логики, но и любые интеллектуальные конструкции, могущие оказать влияние на конечные выводы и увеличить риск недооценки опасности глобальной катастрофы. Некоторые приведённые ошибки могут не приводить в текущих обстоятельствах к каким-либо последствиям, тем не менее, полезно их иметь в виду.

Ошибки, возможные только при анализе и оценке глобальных рисков

Путаница между глобальными катастрофами и просто очень большими катастрофами. Есть тенденция путать глобальные катастрофы, ведущие к уничтожению человечества (обозначаемые в англоязычной литературе термином «existential risks») и любые другие колоссальные катастрофы, которые могут принести огромный ущёрб, отбросить цивилизацию далеко назад и истребить значительную часть человечества. Критерием глобальных катастроф является необратимость. В русском языке пока нет устоявшегося краткого термина для катастроф, ведущих к уничтожению человеческой цивилизации. В этой статье эти катастрофы названы «глобальными катастрофами». Есть ещё термин-калька — экзистенциальные риски. (Подробнее про определение глобальных катаст-

роф и их специфику см. Ник Бостром «Сценарии уничтожения человечества» [Bostrom, 2001].)

Недооценка неочевидных рисков. Глобальные риски делятся на очевидные и неочевидные. Неочевидные риски в некотором смысле гораздо опаснее, потому что неизвестен их объём, их вероятность и в связи с ними ничего не предпринимается. Некоторые неочевидные риски известны только узкому кругу специалистов, которые высказывают диаметрально противоположные мнения в оценке их реальности и вероятности. Эти мнения могут выглядеть для стороннего наблюдателя в равной мере обоснованными, что заставляет его выбирать между мнениями экспертов, или исходя из своих личных предпочтений, или «бросая монетку». Однако неочевидные риски несут вполне реальную угрозу и до того, как научное сообщество окончательно определится с их параметрами. Это заставляет уделять внимание тем областям знаний, в отношении которых ещё много вопросов.

По мере роста наших знаний о природе и могущества техники постоянно росло число известных нам причин возможного уничтожения человечества. Более того, этот рост ускоряется. Поэтому вполне разумно ожидать, что есть важнейшие риски, о которых мы ничего не знаем. И наиболее опасными из них являются те риски, о которых мы не можем ничего узнать, пока события этих рисков не произойдут.

Очевидные риски гораздо удобнее анализировать. Есть огромный объём данных по демографии, военному потенциалу и запасам сырья, которые можно анализировать детально и подробно. Объём этого анализа может заслонять тот факт, что есть другие риски, о которых нам очень мало известно, и которые не годятся для анализа в численной форме, но которые тоже смертельно опасны (например, проблемы с неправильно запрограммированным ИИ).

Нетрудно заметить, что в момент развития аварийной ситуации, например, в авиации, самые страшные последствия имеет именно непонимание пилотами того, что происходит (особенно ошибки в оценке высоты и степени опасности процесса). Наоборот, когда такое понимание имеется, самолёт удаётся спасти часто в совершенно невероятных условиях. И хотя для нас апостериори причины катастрофы очевидны, для самих пилотов они были в тот момент совершено неочевидны.

Глобальные риски нетождественны национальной безопасности. Каждая страна тратит на национальную безопасность больше денег, чем

на глобальную. Однако глобальные риски представляют большую угрозу для каждой страны, чем национальные — просто потому что если погибнет весь мир, то и страна вместе с ним. При этом часто те действия, которые увеличивают безопасность данной страны на текущем этапе, уменьшают всеобщую безопасность. Например, безопасность некой страны возрастает, — во всяком случае, по мнению её руководства — когда она накапливает запасы ядерного и бактериологического оружия, но безопасность всего мира в результате гонки вооружений падает.

Ошибка, связанная с психологизацией проблемы. Издавна существует стереотип сторонника «конца света», толкователя апокалипсиса, — как отверженного обществом индивида, пытающегося своими нелепыми высказываниями повысить свою социальную значимость и компенсировать, таким образом, свои неудачи в финансах и личной жизни. Вне зависимости от истинности или ложности такой интерпретации психологических мотивов людей, это не влияет на степень рисков. Только научные изыскания могут способствовать определению реальной опасности рисков.

Отождествление глобальной катастрофы со смертью всех людей и наоборот. Вымирание человечества не означает гибели всех людей, и наоборот. Легко можно представить себе сценарии, когда большая часть человечества гибнет от некой эпидемии, но один остров уцелеет и за 200 лет восстановит человеческую популяцию и использовавшиеся технологии. Однако если все люди заболеют вирусом, переводящим мозг в состояние непрерывного созерцательного блаженства, то это будет конец цивилизации, хотя огромное большинство людей будет некоторое время ещё живо. Или если — в неком фантастическом сценарии — инопланетяне завоевывают Землю и распродают людей по космическим зоопаркам. Более того, все живущие в настоящий момент люди, если не будет изобретено радикальное средство продления жизни, вымрут к началу XXII в., равно как сейчас вымерли люди, жившие в XIX в. Но мы не рассматриваем это как глобальную катастрофу, потому что сохраняется непрерывность человеческого рода. Настоящая же глобальная катастрофа лишит человечество будущего.

Стереотип восприятия катастроф, который сложился в результате работы СМИ, создаёт ложный образ глобальной катастрофы, что может оказывать подсознательное влияние на оценки. Опыт потребления телевизионных репортажей о катастрофах выработал стереотип, что

конец света нам покажут по CNN. Однако глобальная катастрофа затронет каждого, и смотреть репортажи будет некому. Равно как и показывать.

Ошибки, связанные с тем, что глобальная катастрофа ещё ни разу с нами не происходила. Это имеет несколько последствий:

1) **Пренебрежение неким сценарием как фантастическим** — но глобальная катастрофа не может быть чем-то иным, чем «фантастическим» событием.
2) **Ошибка, возникающая в связи с невозможность осознать тот факт, что никакие события невозможно опознать как глобальную катастрофу ни заранее, ни в процессе — а только апостериори.** Возможно, что никто не узнает, что это на самом деле была глобальная катастрофа. (Однако в сценариях некого медленного вымирания люди могут это осознавать — или заблуждаться на этот счёт. Возможный пример подобного сценария описан в романе Нейджела Шюта «На берегу», где люди медленно вымирают от последствий радиоактивного загрязнения.)
3) **Неприменимость логической операции «индукции» для рассуждений о глобальных катастрофах.** Индукция как логический метод состоит в том предположении, что если некое утверждение верно в моменты 1,2, 3 … N, то оно верно и при N + 1 (или при всех N). Оно не обладает абсолютной логической достоверностью, но даёт хорошие результаты при очень больших N и гладких условиях. Например, все физические законы основаны на конечном количестве экспериментов, т. е. возникли в результате индукции. Однако, индукция как логический метод имеет границы применимости. Этот метод неприменим в ситуациях, когда будущее не похоже на прошлое. В таких случаях мы не можем, на основании того, что нечто было в прошлом всегда, сказать, что так оно будет и в будущем. Однако индукция как логическая процедура применима в вопросах безопасности: с точки зрения обеспечения безопасности трёхразовое периодическое повторение опасного события — говорит о значимости его риска, тогда как с точки зрения доказательства истинности некой закономерности — нет.

Ошибка, состоящая в том, что размышления о глобальных рисках автоматически включают некий архетип «спасителя мира», при этом недооценивается опасность конкуренции между разными группами людей, защищающими разные модели спасения мира. В конце концов,

каждая мировая религия занимается спасением всего человечества, а остальные ей только мешают. Так что борьба спасителей мира между собой может угрожать жизни на Земле. Можно вспомнить анекдот советских времён: «Войны не будет, но будет такая борьба за мир, что от мира камня на камне не останется».

Недооценка глобальных рисков из-за психологических механизмов игнорирования мыслей о собственной смерти. Людей не волнуют глобальные риски, потому что они и так привыкли к неизбежности своей личной смерти в ближайшие десятилетия и выработали устойчивые психологические механизмы защиты от этих мыслей. Наибольший срок реального планирования (а не умозрительных фантазий) можно проследить по долгосрочным реальным инвестициям людей. Отражением которого является покупка дома в ипотеку, пенсионные накопления и воспитание детей — предельный срок этих проектов — 30 лет, за редким исключением, а обычно меньше 20. Однако не факт, что такое планирование на самом деле эффективно, и люди в большинстве своём знают, что жизнь гораздо более непредсказуема. В любом случае, каждый человек имеет некий горизонт событий, происходящее за пределами которого представляет для него чисто умозрительный интерес, а большинство людей считает, что глобальные риски отстоят от нас на многие десятилетия.

Ошибки, связанные с тем, что тот, кто исследует глобальные катастрофы в целом, вынужден полагаться на мнения экспертов в разных областях знания. Часто обнаруживается, что есть множество мнений о какой-либо проблеме, которые выглядят в равной мере аргументировано. А. П. Чехов писал: «Если от болезни предлагается много средств, значит, она неизлечима». В силу этого исследователь глобальных рисков должен быть экспертом по правильному отбору и сопоставлению экспертных мнений. Поскольку это не всегда возможно, велика вероятность неправильного выбора пула экспертов и неправильного понимания ими проблематики глобальных рисков.

Ошибка, связанная с тем, что глобальным рискам, как целому, уделяют меньше внимания, чем рискам отдельных объектов. Глобальные риски должны оцениваться по той же шкале, что и риски всех других составляющих цивилизацию элементов.

Ошибка, связанная с тем, что риск, приемлемый для одного человека или проекта, распространяется на всё человечество. Идеи та-

кого рода: «Человечеству стоит рискнуть на одну сотую процента ради этого нового необычайного результата» являются порочными, потому что так может рассуждать одновременно очень много исследователей и конструкторов, каждый из которых при этом завышает безопасность своего проекта, что в сумме может давать очень высокий риск.

Отсутствие ясного понимания того, к кому, собственно, обращены дискуссии о глобальных рисках. Обращены ли они к гражданам, которые всё равно ничего сделать не могут, к гражданской ответственности учёных, существование которой ещё надо доказать, к правительствам крупных мировых держав или ООН, занятой своими делами, или к неким комиссиям и фондам, специально нацеленным на предотвращение глобальных рисков — чья способность влиять на ситуацию неизвестна. Удручает и отсутствие систематического досье на все риски — с которым все были бы согласны.

Особенность связи теоретического и практического в отношении глобальных рисков. Вопрос о глобальных рисках является теоретическим, поскольку такое событие, как уничтожение всей цивилизации, в ее истории может произойти только один раз, и тем самым прекратить эту историю. И поэтому невозможно проверить ни одну из гипотез экспериментально. Однако должны приниматься практические меры, чтобы глобальных катастроф не случилось. При этом не всегда можно точно сказать благодаря чему был достигнут положительный результат — то что некое событие риска не произошло.

Ошибочное представление о том, что глобальные риски есть что-то отдалённое и не имеющее отношения к ближайшему будущему. В действительности, шанс погибнуть в глобальной катастрофе для молодого человека в текущих исторических условиях, возможно, выше, чем от других причин личной или групповой смертности. Многие факторы глобального риска уже созрели, а другие могут оказаться более зрелыми, чем нам об этом известно (передовые био- и ИИ- исследования).

Легкомысленное отношение к глобальным рискам, возникающее из ошибочного представления о том, что в случае глобальной катастрофы гибель обязательно будет лёгкая и безболезненная — как будто выключили свет. Наоборот, она может быть мучительна и морально (осознание своей вины и столкновение со смертью близких), и физически.

Представление о том, что книги и статьи о глобальных рисках могут значительно изменить ситуацию. Опыт показывает, что даже когда члены правлений компаний, принимавших критически опасные решения, высказывались против — их не слушали. Тем более не стоит ожидать, что эти люди прислушаются или хотя бы будут читать высказывания тех, кто находится за пределами их круга.

Однако и противоположное мнение ошибочно: что глобальные риски либо неизбежны, либо зависят от случайных, неподвластных человеку факторов, **либо зависят от недосягаемых правителей, повлиять на которых невозможно**. Наоборот, циркуляция определённых идей в обществе может создать общественное мнение, которое повлияет на те или иные механизмы принятия решений.

Ложная картина глобальных рисков, создаваемая средствами массовой информации. Часто оказывается, что реальный ущерб непропорционален никаким образом его информационному освещению. Поскольку человек склонен к бессознательному обучению, да и вообще количество утверждений, которые можно воспринять критически, ограниченно, эти идеи создают определённый информационный фон для любых рассуждений о глобальных рисках (наравне с кино и научной фантастикой).

Гордыня исследователя. Занятия анализом глобальных рисков может вызвать у человека ощущение, что он делает самое важное дело во вселенной, а потому является сверхценной личностью. Это может привести в определённых обстоятельствах к тому, что он будет более глухим к новой поступающей информации. Окружающие люди легко будут считывать это состояние личности исследователя, что будет компрометировать тему, которой он занимается. Также не следует забывать закон Паркинсона о том, что каждый человек стремится достичь уровня своей некомпетентности. Глобальный уровень является наивысшим для всех областей знаний.

Интуиция как источник ошибок в мышлении о глобальных рисках. Поскольку глобальные риски относятся к событиям, которые никогда не случались, они контринтуитивны. Интуиция может быть полезна для рождения новых гипотез, но не как способ предпочтения и доказательства. Вера в силу своей интуиции ещё больше способствует ослеплению своими откровениями. Кроме того, интуиция, как проявление бессознательно-

го, может находиться под воздействием неосознаваемых предубеждений, например, скрытого нежелания видеть разрушения и гибель — или наоборот, потребности видеть их там, где их нет.

Использование традиционных научных подходов при исследовании глобальных рисков крайне проблематично. Эксперимент не является способом установления истины о глобальных рисках. В связи с невозможностью эксперимента невозможно объективно измерить, какие ошибки влияют на оценку глобальных рисков. Не может быть статистики по глобальным рискам. Фундаментальная концепция «опровержимости» также неприменима к теориям о глобальных рисках.

Методы, применимые к управлению экономическими и прочими рисками, не применимы для глобальных рисков — их нельзя застраховать, на них невозможно ставить пари: некому и нечего будет выплачивать в случае страхового случая.

Трудности в определении понятия глобального риска в связи с нечёткостью его объекта. Нечёткость относится как к тому, как проводить границы человечества, так и к тому, что именно считать «необратимым повреждением его потенциала». «Пограничный вопрос» касается обезьян-приматов, неродившихся детей, коматозников, преступников, неандертальцев, разумных машин с искусственным интеллектом и других возможных крайних случаев. Важно помнить историческую изменчивость этих границ — ещё пару сотен лет назад дикаря или раба не считали за человека даже образованные люди. Было ли истребление неандертальцев с их точки зрения гибелью человечества? Согласны ли мы, чтобы нас заменили разумные роботы? Не лучше ли смерть, чем насильственное зомбирование во враждебную религию? Дело даже не в самих этих вопросах, а в том, что ответ на них зависит от человеческого произвола, что приводит к тому, что одни группы людей буду считать «глобальной катастрофой» то, что другие будут готовы приветствовать. Это создаёт возможности для опасной конфронтации.

Структурный кризис — это не тот кризис, который предусмотрен нашими моделями. Это кризис самих моделей (Панов, LTCM). Это, в частности, значит, что те модели, которыми мы привыкли оперировать в отношении личных, промышленных и прочих катастроф часто оказываются непригодны для решения задач связанных с глобальными катастрофами.

Психологическая особенность человека, называемая «пренебрежение масштабом». Спасение жизни одного ребёнка, миллиона человек, миллиарда или ста миллиардов вызывает почти одинаковое побуждение действовать, в том числе выражаемое в готовности потратить деньги (Юдковский, 2006).

Преувеличение прогностической ценности экстраполяции. Иначе говоря, потенциальная ошибочность идеи о том, что «кривая вывезет». Для некоторых закон Мура об удвоении числа транзисторов на процессоре каждые два года стал религией. Увы, весь опыт футурологии говорит о том, что экстраполяция кривых годится только для краткосрочных прогнозов. В более прикладной футурологии, которой является биржевая аналитика, наработан огромный аппарат анализа поведения кривых, превосходящий многократно линейную экстраполяцию, как если бы эти кривые были самостоятельными живыми организмами. В частности, там развито понимание, что быстрый рост кривой может означать близкий разворот тенденции, «отскок» или «свечу». И, тем не менее, даже биржевая аналитика кривых не даёт высокоточных результатов без привлечения «фундамента» — анализа реальной экономики. Особенно за счёт эффекта обратной связи между предсказаниями и самими будущими событиями. Количество ошибок в футурологии, основанных на экстраполяции кривых, огромно. Начиная с того, что уровень навоза в Лондоне достигнет уровня крыш, и вплоть до прогнозов освоения Марса к концу XX в. на заре успехов космонавтики. В отношении глобальных рисков есть определённые ожидания, что прогресс в области техники сам собой нас приведёт в золотой век, где глобальных рисков не будет. Действительно, есть надежды, что будущие технологии искусственного интеллекта станут основой для гораздо более быстрого и эффективного решения проблем. Однако, глобальные риски в значительной их части никуда не денутся.

Ошибочное представление о том, что люди в целом не хотят катастрофы и конца света. Назаретян [20] описывает базовую потребность людей в сильных эмоциях, которая побуждает их нарушать скоростной режим движения машин, вести войны, заводить любовниц, короче, находить приключения. При этом люди всегда «рационализируют» эти свои иррациональные потребности, объясняя их каждый раз тем, что так на самом деле нужно. Иначе говоря, нельзя недооценивать скуку. (Типичный пример тому — то, что можно было бы ограничить число автомобильных аварий в разы, введя физическое ограничение скорости машин да 50 км/час, но общество в этом не заинтересовано.)

Смутность представлений о том, что именно является «поражающим фактором» в разных сценариях глобального риска. Упор делается в основном на непосредственные причины, а не на ситуации их возникновения.

Ошибки, связанные с разными горизонтами восприятия возможного будущего у разных людей. Выделяют пять уровней «шока будущего». Само понятие введено футурологом Тофлером. Эти уровни описывают не реальные границы возможного будущего, которые нам пока неизвестны, а психологические границы восприятия, разные у разных людей. Каждому уровню возможного будущего соответствуют свои глобальные риски — и способы противостояния. При этом все эти варианты будущего относятся их сторонниками к XXI веку. Те, кто продвинулся очень далеко по шкале шоков будущего могут недооценивать традиционные опасности.

Шок 0 уровня — уровень обычных технологий, используемых сейчас в быту. (Уровни катастрофы: ядерная война, исчерпание ресурсов.)

Шок 1 уровня — уровень технологий, предлагающихся в продвинутых журналах и на компьютерных выставках. (Биологическая война и применение боевых роботов.)

Шок 2 уровня — технологии, описываемые к классической научной фантастике середины XX в. (Отклонение астероидов в сторону Земли, вторжение инопланетян.)

Шок 3 уровня — сверхтехнологии, которые появились на горизонте только в конце XX в.: нанотехнологии (умная пыль), ИИ, равный человеческому или превосходящий его, загрузка сознания в компьютер, полная перестройка человеческого тела. (Катастрофы: серая слизь, сверхчеловеческий ИИ, перерабатывающий всё земное вещество в роботов, супервирусы, изменяющие поведение людей.)

Шок 4 уровня — Сингулярность — гипотетический момент в будущем, связанный бесконечным ускорением человеческого прогресса и неким качественным переходом и сменой модели развития. (Риски: непредсказуемы.)

Риски ошибок, связанные с Шоком будущего, состоят в том, что каждый человек, моделирующий будущее, имеет разный горизонт представления о возможном и невозможном, определяемый скорее его психологическим комфортом, чем точными знаниями. Чем старше человек, тем

труднее ему принять новое. Наоборот, возможна и ситуация «ослепления будущим», когда риски невероятных катастроф затмят в глазах человека обычные риски. При этом свои риски глобальной катастрофы есть на каждом уровне. Катастрофа в виде ядерной войны понятнее, чем псевдодружественный ИИ.

Представление о том, что глобальная катастрофа будет вызвана какой-то одной причиной. Обычно о глобальных катастрофах думают, как об однократной массовой гибели людей, вызванной или вирусом, или падением астероида, или ядерной войной. Однако существуют способы самоорганизации опасных возможностей, которые создают системный эффект. Например, система, ослабленная одним событием, может быть легко повержена другим. Или, например, две медленно текущие болезни, соединяясь, могут давать быстротекущую — как СПИД и туберкулёз у человека. Возможны разные сценарии конвергенции, например, нанотех упростит создание ядерных бомб, ИИ упростит создание нанотеха, а нанотех позволит узнать тайны мозга, что приблизит создание ИИ. Так конвергенция рисков идёт параллельно конвергенции ключевых современных технологий, называемых NBIC: nano-bio-iniligence-cogno, т. е. нанотехнологий, биотехнологий, систем искусственного интеллекта и науки о мышлении и мозге.

Недооценка системных факторов глобального риска. Системные факторы — это не отдельные события, вроде внезапного появления супервируса, а некие свойства, которые относятся ко всей системе. Например, противоречие между природой современной цивилизации, основанной на непрерывной экспансии, и конечностью любых ресурсов. Оно не находится в каком-то одном месте, и не зависит ни от одного конкретного ресурса или человеческой организации. Существуют самовоспроизводящиеся кризисные ситуации, которые в каждый момент времени вовлекают в себя всё большое число участников сообщества — но не зависят от поведения ни одного из них и не имеют центра.

Вот несколько примеров самовоспроизводящихся структурных кризисов в качестве иллюстрации (примеры в скобках в определенной степени условны, так как возможны и другие интерпретации тех же событий)

- Превышение критического числа хищников в экосистеме (Остров Пасхи).
- Разрастающаяся трещина в отношениях (конфликт сверхдержав).

- Самовоспроизводящаяся паранойя (репрессии при режиме Полпота в Кампучии и при режиме Сталина в СССР).
- Самовоспроизводящаяся дезорганизация (парад суверенитетов в СССР в 1991 г.).
- Самоподдерживающаяся моральная деградация (крах Римской империи).
- Скручивающаяся спираль экономики (Великая депрессия).
- Эффект домино (наркоман, который подсаживает других, чтобы достать деньги на новую дозу).
- «Естественный» отбор краткосрочных выгод вместо долгосрочных (Маркс: «злые капиталисты вытесняют добрых»).

Очевидно, что конвергенция глобальных рисков из предыдущего пункта и системная природа рисков также могут определённым образом создавать систему ещё более высокой сложности и катастрофичности.

Недооценка предкризисных событий как элементов глобальной катастрофы. Если в результате некоторых событий вероятность глобальной катастрофы возросла (иначе говоря, выросла уязвимость человечества к катастрофе), то это событие само можно рассматривать как часть глобальной катастрофы. Например, если в результате ядерной войны выживут отдельные группы людей, то они — немногочисленные и лишённые технологий — окажутся гораздо уязвимее к любым другим факторам риска. Это возвращает значение тем факторам, которые обычно называются «глобальными рисками» — например, падение астероида, размером с Апофис (400 м) само по себе не может истребить человечество, так как взрыв составит только 800 мегатонн, что сопоставимо с взрывом вулкана Санторин в древней Греции, погубившем остров Крит, и только в 4 раза сильнее взрыва вулкана Кракатау в 1883 г., оцениваемому в 200 мегатонн тротилового эквивалента. Однако в связи с тем, что связность современной цивилизации значительно возросла, возросла и роль отдалённых — экономических и структурных последствий — разных катастроф. Огромное цунами от падения Апофиса привело бы к прекращению торговли в тихоокеанском регионе и всеобщему экономическому кризису, чреватому переходом к военным режимам повсеместно — с соответствующим нарастанием необратимых последствий в виде спада производительных сил и эффективности мировой экономики.

Ошибки, возможные относительно оценки любых рисков, применительно к глобальным рискам

Основным источником любых рукотворных катастроф является человеческий фактор, а основной причиной человеческих ошибок является сверхуверенность. Сверхуверенность означает повышенную убеждённость в правильности сложившейся картины мира и невозможность изменить сложившиеся представления даже при поступлении новых фактов, говорящих о их ее ошибочности. Иначе говоря, сверхуверенность означает неспособность человека предположить, что он в настоящий момент ошибается. Сверхуверенность закреплена несколькими механизмами в человеческом сознании, и вероятно, имела эволюционное значение — поэтому обнаружить её в себе и искоренить её очень трудно. Человек, демонстрирующий большую уверенность, мог претендовать на большую власть в обществе.

И, вероятно, само устройство человеческого мышления противоречит идее рассмотрения множества равновероятных сценариев будущего: гораздо привычнее думать о том, что нечто или есть, или его нет. Наконец, однажды сделанный выбор в сторону одной интерпретации становится фильтром, который не пропускает информацию не подтверждающую сложившиеся взгляды.

Чрезмерное внимание к медленно развивающимся процессам и недооценка быстрых. Медленный процессы удобнее анализировать, и по ним накоплено больше данных. Однако системы легче адаптируются к медленным изменениям и гибнут часто от быстрых. Катастрофы опаснее угасания. При этом медленное угасание делает систему уязвимой к быстрым катастрофам.

Споры не рождают истину о глобальных рисках. Дискуссии между людьми обычно приводят к поляризации мнений. То есть человек, который имел в начале две гипотезы, которым приписывал разные вероятности, редуцирует свою позицию до одной гипотезы, противоположной гипотезе оппонента. Таким образом, он сужает своё представление о возможном будущем. Подробнее см. в [Юдковский, 2007].

Обнаружение ошибок в рассуждении о возможности некой конкретной катастрофы не является способом укрепления безопасности. Есть два вида рассуждений — доказательства безопасности некой системы

или доказательства существующих опасностей. Эти рассуждения неравноценны логически — в первом случае речь идёт обо всех возможных случаях, тогда как во втором — хотя бы об одном случае. Чтобы опровергнуть всеобщее утверждение, достаточно одного контрпримера. Однако опровержение одного контрпримера не прибавляет истинности всеобщему утверждению.

Например: для того, чтобы доказать опасность некоего самолёта, достаточно указать на то, что в некоторых экспериментах металл обшивки проявил склонность к эффекту «усталости металла».

Однако для того, чтобы доказать безопасность самолёта, совершенно недостаточно обнаружить некорректность в проведении этих экспериментов по измерению усталости металла. Вместо этого необходимо доказать, что выбранный материал действительно выдержит данный режим нагрузок.

Иначе говоря, если направить все интеллектуальные усилия на опровержение отдельных катастрофических сценариев, пренебрегая другими — то суммарная безопасность системы понизится.

Ошибочное представление о том, что когда проблема назреет, тогда к ней можно начать готовиться. Наиболее серьёзные проблемы возникают внезапно. Чем серьёзнее проблема, тем больше её энергия и — возможно — темп её развития. И тем труднее к ней готовиться. Глобальные катастрофы — это мощные проблемы, поэтому они могут развиваться слишком быстро, чтобы к ним успеть подготовиться в процессе. Кроме того, у нас нет опыта, который позволил бы определить предвестников глобальной катастрофы заранее. Например, несчастные случаи развиваются внезапно. «Знал бы, где упаду — соломку бы постелил».

Более конкретизированные риски воспринимаются как более опасные, чем описанные в общих словах. Например, «мятеж на ядерной подводной лодке» выглядит более устрашающе, чем «крупная морская катастрофа». «С точки зрения теории вероятностей, добавление дополнительной детали к истории делает её менее вероятной … Но с точки зрения человеческой психологии *добавление каждой новой детали делает историю всё более достоверной.»* [Yudkowsky, 2007] Подробнее см. в статье Юдковского, где описаны конкретные психологические эксперименты, подтверждающие эту систематическую ошибку.

Утрата чувствительности общества к предупреждениям

Широко распространенное представление о том, что мышление о глобальных рисках — источник пессимизма. Это приводит к тому, что людей, думающих о «конце света», осуждают — а значит, и отрицают их идеи. По минному полю надо идти осознано — танцевать на нём с закрытыми глазами — это не оптимизм.

«Теории заговора» как препятствие для научного анализа глобальных рисков. Циркулирующие в обществе разнообразные «теории заговоров», вроде новой хронологии Фоменко, представляют особую опасность. Большинство (если не все) из них, ложно, а их предсказания почти никогда не сбываются. Часто «теории заговоров» тоже предсказывают некие риски. Но они структурно отличаются от научного анализа рисков. Теория заговора обычно утверждает, что человечеству угрожает только один риск, и этот риск реализуется конкретным образом в конкретный момент времени: Например, «доллар рухнет осенью 2007». Как правило, автор также знает рецепт, как с этим риском бороться.

«Теории заговора» вредны для предсказания будущего, так как сужают представление о множестве будущих возможностей. При этом они предполагают сверхуверенность в собственных прогностических способностях. Хорошее предсказание будущего не предсказывает конкретные факты, а описывает пространство возможных сценариев. И на основании этого знания можно выделить узловые точки этого пространства и решать связанные с ними проблемы.

Более того, такие «предсказания» подрывают доверие к здравым идеям, лежащим в их основе, например, о том, что крупный теракт может ослабить доллар и вызвать цепную реакцию обвала. И что Бен Ладен это тоже понимает, и на это, возможно, рассчитывает. «Теории заговора» всегда подразумевают, что есть некие ОНИ, которые с нами что-то делают, скрывают и т. д. Это подрывает наше осознание своей ответственности за происходящее в мире, и что не менее важно, отвергает роль случайности, как важного фактора катастрофического процесса. Кроме того, «теории заговора» неспособны стыковаться друг с другом, формулируя пространство возможностей. И ни одна теория заговора не признаёт себя в качестве

таковой. Эти теории распространяются в обществе как мемы, самокопирующиеся информационные единицы.

Вместе с тем, из того, что сам принцип теории заговоров скомпрометирован и большинство из них ложно, не следует, что некоторые, тем не менее, не могут оказаться правдой. «Даже если вы не можете поймать чёрного кота в чёрной комнате — это ещё не значит, что его там нет».

Ошибки, связанные с путаницей краткосрочных, среднесрочных и долгосрочных прогнозов. Краткосрочный прогноз учитывает текущее состояние системы, к таковым относится большинство обсуждений на тему политики. Среднесрочный учитывает возможности системы и текущие тенденции. Долгосрочный учитывает только развитие возможностей. Поясню это следующим примером:

Допустим, у нас есть корабль с порохом, по которому ходят матросы и курят махорку. Краткосрочно можно рассуждать так: одни матросы высоко на рее, а другие спят, поэтому сегодня взрыва не будет.

Но в среднесрочной перспективе важно только количество пороха и количество курящих матросов, которые определяют вероятность взрыва, потому что рано или поздно какой-нибудь курящий матрос окажется в неправильном месте. А в дальнесрочной перспективе в счёт идёт только количество пороха, а уж огонь как-нибудь да найдётся.

Точно так же и с угрозой ядерной войны. Когда мы обсуждаем её вероятность в ближайшие два месяца, для нас имеют значение телодвижения мировых держав. Когда мы говорим о ближайших пяти годах, в счёт идёт количество ядерных держав и ракет. Когда мы говорим о перспективе на десятки лет, в счёт идёт только количество наработанного плутония.

При этом в разных областях знания временной масштаб краткосрочности прогноза может различаться. Например, в области угледобычи — 25 лет — это краткосрочный прогноз. А в области производства микропороцессоров — 1 год.

Специфика человеческой эмоции страха. Способность бояться включается в ответ на конкретный стимул в конкретной ситуации. Наш орган страха не предназначен для оценки отдалённых рисков. Это выражено в русской пословице: «Пока гром не грянет, мужик не перекрестится».

Эффект смещения внимания. Чем больше некий человек уделяет внимания одной возможной причине глобальной катастрофы, тем меньше он уделяет другим, и в результате его знания приобретают определённый сдвиг в сторону его специализации.

Склонность людей бороться с подобными опасностями, какие уже были в прошлом. Например, было цунами 2004 г. — теперь все стали строить системы предупреждений. А в следующий раз это будет не цунами. При этом с течением временем тревога людей убывает, а вероятность повторного сильного землетрясения (но не афтершока) — возрастает.

Усталость от ожидания катастрофы. Типична ошибка, состоящая в том, что после того, как некоторая катастрофа случится, все начинают ожидать повторения в ближайшем будущем второй точно такой же, и после того, как это ожидание не исполняется, переводят эту катастрофу в разряд «это было давно и неправда». Так было после теракта 11 сентября. Сначала все ждали повторных атак на небоскрёбы, и строительство небоскрёбов в мире затормозилось. Теперь же все об этом забыли, и строительство небоскрёбов идёт ударными темпами. Это противоречит тому, что в реальности катастрофы такого масштаба могут происходить с периодичностью во много лет, и поэтому именно после длительного промежутка времени их вероятность реально возрастает.

Экспертные оценки, не основанные на строгих вычислениях, не могут служить в качестве меры реальной вероятности. (В отличие от ситуации на фондовых рынках, где среднестатистическая оценка лучших экспертов используется для предсказания будущего результата. Увы — или к счастью — мы не можем калибровать и отбирать наших экспертов по глобальным рискам по количеству угаданных ими катастроф.)

Обнаружена следующая статистика в экспериментах по предсказанию: «Только 73 % ответов, на которые сделали ставки 100 : 1, были верны (вместо 99,1 %). Точность возросла до 81 % при ставках 1000 : 1 и до 87 % при 10.000 : 1. Для ответов, на которые ставили 1.000.000 : 1, точность составляла 90 %, т. е. соответствующий уровень доверия должен был бы порождать ставки 9 : 1. В итоге, испытуемые часто ошибались даже при высочайших уровнях ставок. Более того, они были склонны делать очень высокие ставки. Более чем половина их ставок была более чем 50 : 1». [Yudkowsky, 2007]

Подобные уровни ошибок были обнаружены и у экспертов. Hynes и Vanmarke (1976) опросили семь всемирно известных геотехников на предмет высоты дамбы, которая вызовет разрушение фундамента из глинистых пород, и попросили оценить интервал 50 % уверенности вокруг этой оценки. Оказалось, что ни один из предложенных интервалов не включал в себя правильную высоту». [Yudkowsky, 2007]

Причиной этой ошибки является «сверхуверенность экспертов» — например, потому что эксперт боится потерять свой статус эксперта, если будет слишком сомневаться в своих мнениях.

Игнорирование какого-либо из рисков по причине его незначительности по мнению эксперта. «Этого не может быть, потому что не может быть никогда». Эта теория ненаучна и не доказана, поэтому мы не будем её рассматривать.

Ограниченность числа свободных регистров в уме человека и модель мышления, отражающаяся в каждом предложении: субъект — объект — действие. Это заставляет человека концентрироваться на одних аспектах проблемы, вроде того, нападёт ли А на Б, уменьшая при этом, — погружая в тень внимания, — другие аспекты. Ни один человек не может охватить все мировые проблемы в своём уме, чтобы ранжировать их по степени их опасности и приоритетности. Вряд ли это может и одна какая-либо организация.

Раскол футурологии по разным дисциплинам, как если бы эти процессы происходили независимо. Есть несколько направлений мышления о будущем, и они имеют склонность странным образом не пересекаться, как будто речь идёт о разных мирах:

- прогнозы о «Сингулярности» (суперкомпьютеры, биотехнологии, и нанороботы);.
- прогнозы о системном кризисе в экономике, геополитике и войнах;
- прогнозы в духе традиционной футурологии о демографии, ресурсах, потеплении и т. д.

Отдельной строкой: большие катастрофы: астероиды, супервулканы, сверхвспышки на солнце, переполюсовка магнитного поля. Плюс религиозные сценарии и фантастические сценарии.

Ситуация, когда вслед за меньшей проблемой следует большая, но мы неспособны этого заметить. («Беда не приходит одна»)

Причины этого могут быть заключены в следующем:

1) Внимание в момент столкновения с первой проблемой полностью отвлекается и разрушается. Например, попав в маленькую аварию, водитель начинает ходить вокруг машины, и тут в его сбивает другая машина.
2) Переход в состояние аффекта. Так при пожаре может возникнуть желание выпрыгнуть в окно.

3) Нередко в процессе исправления небольшой ошибки совершается еще большая. Например, когда мелкий воришка стреляет в полицейского, чтобы скрыться.
4) Непонимание того, что первая авария создаёт неочевидную цепочку причин и следствий, которая может повлечь за собой цепочку еще боле опасных событий. Например, грипп чреват воспалением лёгких, при неправильном лечении. То есть первая неприятность постепенно ослабляет сопротивляемость организма к более быстрым и внезапным переменам.
5) Непонимание того, что обе аварии могут быть вызваны некой неочевидной общей причиной. Например, что-то обвалилось, человек пошёл посмотреть — что, и тут оно обвалилось целиком.
6) Эйфория от преодоления первой катастрофы может заставить потерять благоразумие. (Например, человек может так рваться выйти из больницы пораньше, что у него разойдутся швы.)

Эффект избирательности внимания, сосредоточенного на катастрофах. Часто у людей, следящих за некими предсказаниями, например в экономике, возникает вопрос: «Почему то, что должно вот-вот рухнуть, всё не рушиться и не рушится?» Вероятно, мы имеем дело со специфической ошибкой в оценке рисков. Когда мы замечаем трещины в фундаменте, мы говорим себе: «О! Оно же вот-вот рухнет» и начинаем искать другие трещины. Разумеется, мы их находим, и нам не трудно связать их в умозрительную сеть. Но, занимаясь поискам трещин, мы перестаём смотреть на опоры. Наше внимание становится избирательным, нам хочется подтвердить свою гипотезу. Мы попадаем в порочный круг избирательного накопления информации только об одном аспекте неустойчивости системы, игнорируя причины её устойчивости, а также другие риски, связанные с этой системой. Завышение некоторых рисков, в конечном счёте, приводит тоже к их недооценке, поскольку общество приобретает иммунитет к негативным прогнозам и утрачивает доверие к экспертам. Например, станция предупреждения о цунами на Гавайях оказалась перед дилеммой: если предупредить население о риске цунами, то в следующий раз ей не поверят, а если не предупредить — возможно, что именно в этот раз цунами окажется опасным. Таиландская служба предупреждения в 2004 г. решила не предупреждать людей о цунами, боясь напугать людей.

Подсознательное желание катастрофы. Стремление эксперта по рискам доказать правоту своих прогнозов вызывает у него неосознанное

желание того, чтобы прогнозируемая катастрофа таки случилась. Это подталкивает его или преувеличить предвестники приближающейся катастрофы, или даже попустительствовать тем событиям, которые могут к ней привести. Люди также могут хотеть катастроф от скуки или в силу мазохистского механизма «негативного наслаждения».

Использование сообщений о рисках для привлечения внимания к себе, выбивания денег и повышения своего социального статуса. Этот тип поведения можно назвать «Синдром Скарамеллы», в честь итальянского мошенника, выдававшего себя за эксперта по вопросам безопасности. В крайне остром случае человек выдумывает некие риски, потому что знает, что общество и масс-медиа на них резко реагируют. Эта модель поведения опасна тем, что из общего контекста выдёргивается несколько самых зрелищных рисков, а не менее опасные, но не столь завлекательно звучащие риски затушёвываются. Кроме того, у общества возникает привыкание к сообщениям о рисках, как в сказке о мальчике, который кричал «Волк, Волк!», а волка не было. Когда же волк пришёл на самом деле, никто мальчику не поверил. Более того, возникает общественная аллергия на сообщения о рисках, и все сообщения начинают объясняться в терминах пиара и деления денег.

Использование темы глобальных рисков в качестве сюжета для развлекательных масс-медиа. Выделение адреналина в критических ситуациях по-своему приятно, и небольшой укол его можно получить, посмотрев документальный фильм-катасрофу. Это приводит к тому, что разговоры о рисках начинают восприниматься как нечто несерьёзное, не имеющее отношения к реальности и проблемам, даже как нечто приятное и желанное.

Ошибка «хорошей истории». Описана у Бострома [Bostrom, 2001] Регулярное потребление продуктов масс-медиа подсознательно формирует модель риска, который назревает, угрожает, интересно развивается, но затем зрелищно побеждается, вся игра идёт почти на равных. Реальные риски не обязаны соответствовать этой модели. Юдковский называет это «логической ошибкой генерализации на основании художественного вымысла». [Yudkowsky, 2007]

Даже если мы стараемся избегать воздействия художественных произведений, фильм «Терминатор» сидит у нас в подсознании, создавая, например, ошибочное представление, что проблемы с Искусственным Интеллектом — это обязательно война с роботами.

Идеи о противостоянии глобальным рискам с помощью организации единомышленников, связанных общей целью — благо человечества. Потому что всегда, когда есть «мы», есть и «они». Потому что у любой организации есть самостоятельная групповая динамика, направленная на укрепление и выживание этой организации. Потому что у любой организации есть конкурент. Потому что внутри организации запускается групповая динамика стада-племени, побуждающая бороться за власть и реализовывать другие скрытые цели.

Секретность как источник ошибок в управлении рисками. Исследования по безопасности, ведущиеся в секрете, утрачивают возможность получать обратную связь от потребителей этой информации и в итоге могут содержать больше ошибок, чем открытые источники. Засекречивание результатов неких испытаний и катастроф обесценивает их назначение для предотвращения последующих катастроф.

Сверхкритическая интеллектуальная установка мешает обнаруживать опасные катастрофические сценарии. Сверхкритичность препятствует начальной фазе мозгового штурма, на которой набирается банк возможных идей. Поскольку безопасности часто угрожают невероятные стечения обстоятельств, то именно странные идеи могут быть полезными.

Идея о том, что безопасность чего-либо можно доказать теоретически. Однако единственный реальный критерий — практика. История знает массу примеров, когда приборы или проекты, теоретически имевшие огромную безопасность, рушились из-за непредусмотренных сценариев. Например, фонд LTCM или самолёт Конкорд.

Недооценка человеческого фактора. От 50 до 80 % катастроф происходят из-за ошибок операторов, пилотов и других людей, осуществляющих непосредственное управление системой. Ещё значительная доля катастрофических человеческих ошибок приходится на техническое обслуживание, предполётную подготовку и ошибки при конструировании. Источником большинства ошибок является чрезмерная уверенность в своих знаниях и способностях, приводящая к ошибочной картине развития ситуации. Даже сверхнадёжную систему можно привести в критическое состояние определённой последовательностью команд. Человек достаточно умён, чтобы обойти любую «защиту от дурака» и натворить глупостей.

Ошибка, связанная со склонностью людей в большей мере учитывать широко известные факты. Это приводит к тому, что одни

риски переоцениваются, и уже в силу этого другие риски недооцениваются.

Анализ глобальных рисков не есть создание прогнозов. Прогноз даёт конкретные данные о времени и месте. Но такие точные попадания очень редки и скорее случайны. Более того, прогноз и анализ рисков требует разных реакций. Неудачные прогнозы компрометируют свою тему и людей, которые их дают. Но некоторые люди дают очень много прогнозов, надеясь, что хоть один попадёт в цель и человек прославится. Например, анализ рисков в авиации требует усовершенствования разных механизмов самолёта и организации полетов, а прогноз об авиакатастрофе предполагает, что люди откажутся от рейса в данный день.

Эффект знания задним числом приводит к тому, что люди восклицают: «Я знал это с самого начала», и в силу этого переоценивают свои прогностические способности. Также в силу этого они ждут, что другие люди могли бы легко догадаться о том, что нам уже известно. В отношении глобальных рисков у нас нет никакого знания задним числом. А в отношении многих других обычных рисков есть. Это приводит к тому, что нам кажется, что глобальные риски так же легко оценить, как уже известные нам риски. Иначе говоря, эффект знания задним числом в отношении глобальных рисков приводит к их недооценке.

Эффект настройки на источники информации. Читая литературу, человек в каком-то смысле становится рупором идей, которые в него вкладывает автор. Это позволяет ему сходу отвергать концепции других людей. В силу этого он становится глухим к новой информации, и его эффективность в анализе рисков падает. Ощущение собственной правоты, образованности, навыки ведения споров — всё это усиливает «глухоту» человека. Поскольку глобальные риски — есть вопрос в первую очередь теоретический (ведь мы не хотим экспериментальной проверки), то теоретические разногласия имеют тенденцию в нём проявляться особенно ярко.

Частой ошибкой исследователей рисков является принятие малого процесса за начало большой катастрофы. Например, изменение курса доллара на несколько процентов воспринимается как предвестник глобального краха американской валюты. Это приводит к преждевременным высказываниям прогностического триумфа в духе: «ну вот, я же говорил!» — что затем подрывает веру, в первую очередь, свою собственную, в возможность катастрофы.

Часто, когда некая катастрофа уже произошла, более простое её объяснение подменяет более сложное. На выяснение же более сложного уходят годы анализа, например, авиакатастрофы. Это более сложное объяснение не доходит до широкой публики и остаётся в качестве некоторого информационного фона. Чем дольше не найдено точное определение причин аварии, тем дольше невозможно защититься от аварии подобного рода. Когда речь идёт о быстрых процессах, такое отставание понимания может стать критическим.

Есть класс людей, которые придумывают апокалипсические сценарии, чтобы привлечь внимание к своим безумным проектам и добиться их финансирования. Даже если 99 % этих людей явно не правы, выдвигаемые ими гипотезы, вероятно, следует принимать к сведению, так как ставки в игре слишком велики, и неизвестные физические эффекты могут угрожать нам и до того, как их официально подтвердит наука.

Стремление людей установить некий приемлемые для них уровень риска. У каждого человека есть представление о норме риска. Поэтому, например, водители более безопасных машин предпочитают более опасный стиль езды, что сглаживает в целом эффект безопасности машины. Как бы ни была безопасна система, человек стремится довести её до своей нормы риска.

Эффект «сверхуверенности молодого профессионала». Он возникает у водителей и пилотов на определённом этапе обучения, когда они перестают бояться и начинают чувствовать, что уже всё могут. Переоценивая свои способности, они попадают в аварии.

Ощущение неуязвимости. Это усугубляется эффектом избранности наблюдателя, который состоит в том, что, например, отвоевавшие определённый срок без ранений солдаты начинают чувствовать свою неуязвимость, и всё более и более повышать свою норму риска. Это же происходит с цивилизацией — чем дольше не было, например, атомной войны, тем в большей мере кажется, что она вообще невозможна и тем более рискованную политику можно проводить.

Переоценка собственных профессиональных навыков. Если человек сравнивает то, что он знает, с тем, что он знает, он всегда получает 100 %. Это может создать у него иллюзию, что он знает всё. Тогда как адекватная оценка возможна, только если человек сравнивает то, что он знает, с тем, что он не знает. В этом случае он получает честную границу своих

знаний. Поскольку глобальные риски охватывают все сферы знаний — от биологии до астрофизики и от психологии до политики, то чтобы получить адекватную картинку ситуации, любой специалист вынужден выйти за пределы своих знаний. Альбер Камю: «Гений — это ум, осознавший свои пределы». Поскольку чувствовать себя профессионалом приятно, человек может испытывать склонность к преувеличению своих способностей. Это будет мешать ему проконсультироваться у специалистов по существенным вопросам. Стереотип «спасителя мира» как героя-одиночки, который способен на всё, может помешать ему скооперироваться с другими исследователями и сделать свой ценный вклад. В равное мере и представление об ордене «джедаев», тайно спасающих мир, может быть некорректным и целиком заимствованным из развлекательного кино.

Есть ряд ситуаций, когда меры по предотвращению небольшой катастрофы готовят ещё большую катастрофу. Например, в Йеллоустоунском парке так успешно боролись с пожарами, что в лесу скопилось очень много сухих деревьев и в результате произошёл колоссальный пожар, справиться с которым было почти невозможно.

Утомление исследователя. Энтузиазм отдельных людей движется волнами, и в силу этого человек, который, допустим, начал выпускать некий бюллетень, может, утратив энтузиазм, начать выпускать его всё реже, что с точки зрения стороннего наблюдателя будет означать снижение интенсивности событий в этой области. Тем более работа исследователя глобальных рисков неблагодарна — он никогда не увидит реализации своих пророчеств, кроме самых крайних сценариев. И у него никогда не будет уверенности, что ему на самом деле удалось что-то предотвратить. Только в кино спаситель мира получает за свою работу благодарность всего человечества и любовь красивой актрисы. Так, Черчилль проиграл выборы сразу после войны, хотя он верил, что заслужил переизбрания. Чтобы избежать этого эффекта, на американском флоте во время Второй Мировой войны применяли регулярную ротацию высшего состава — одна смена воевала, а другая отдыхала на берегу.

Страх потери социального статуса исследователями, который приводит к тому, что они не касаются некоторых тем. В нашем обществе есть ряд тем, интерес к которым воспринимается как симптом определённого рода неполноценности. Люди, интересующиеся этими вопросами, автоматически считаются (или даже выдавливаются в соответствующие «экологически ниши») второсортными, сумасшедшими, кло-

унами и маргиналами. И другие исследователи даже могут стремиться избегать контакта с такими людьми и чтения их исследований. Темы, которые заклеймены, это: НЛО, телепатия и прочая парапсихология, сомнение в реальности мира. Однако важно отметить, что если хотя бы одно сообщение об НЛО истинно и не объяснимо, то это требует пересмотра всей имеющейся картины мира, и не может не влиять на вопросы безопасности. Более того, те исследователи, которые потеряли свой статус, проявив интерес к НЛО или чему-то подобному, утратили вместе с этим и возможность доносить свои мысли до представителей власти. Военные исследования в этой области настолько засекречены, что неизвестно, имеются ли такие исследования вообще, и соответственно, в какой мере можно доверять людям, говорящим от имени этих исследований. Иначе говоря, секретность настолько инкапсулирует некую исследовательскую организацию, что она перестаёт существовать для внешнего мира, как чёрная дыра, — особенно в тех случаях, когда даже высшее руководство страны может не знать о ней. (Характерен пример с канцлером Меркель, которой отказывались объяснять, что за люди ходят по резиденции, пока она это категорически не потребовала — это оказались сотрудники службы безопасности.)

Количество внимания, которое общество может уделить рискам, ограничено. Поэтому преувеличение некоторых рисков не менее опасно, чем умалчивание о других, так как съедает то количество внимания (и ресурсов), которые можно потратить на более опасные риски. Кроме того, оно создаёт ложную успокоенность у человека, которому кажется, что он сделал достаточный вклад в спасение Земли, например, заправив свой автомобиль спиртом.

Пренебрежение экономикой. Такие выражения, как «деньги — это только бумажки», или «банковские вклады — это только нолики в компьютерах» могут быть отражением широко распространённого мнения, что экономика не так важна, как, скажем, война или некие более зрелищные катастрофы. Однако экономика — это материальное воплощение структурности всех происходящих на Земле процессов. Для понимания роли экономики важно отметить, что кризис 1929 г. нанёс США ущерб в 2 раза больший, чем Вторая мировая война, а крах СССР произошёл не в результате прямой агрессии, а результате структурно-экономического кризиса. Крестьянская община в России исчезла не от вирусов и войн, а от тракторов и массового переезда в города. Даже вымирание динозавров и другие

крупные вымирания многие биологи связывают не с космической катастрофой, а с изменением условий конкуренции между видами.

Все риски имеют стоимостное выражение. Экономические последствия даже небольших катастроф могут иметь огромное стоимостное выражение. Теракты 11 сентября нанесли ущерб американской экономике в 100 млрд долл., и возможно, гораздо больше, если учесть потенциальный ущерб от политики низкой процентной ставки (пузырь на рынке недвижимости), а также триллионы долларов, потраченные на войну в Ираке. При этом цена разрушенных зданий составляла только несколько миллиардов долларов.

Итак, даже небольшие аварии могу приводить к огромному ущербу и утрате стабильности экономики, а крах экономики сделает систему менее устойчивой и более уязвимой к ещё большим катастрофам. Это может привести к положительной обратной связи, т. е. к самоусиливающемуся катастрофическому процессу.

По мере глобализации экономики, всё больше возрастает возможность всепланетного системного кризиса. Конечно, трудно поверить, что мир погибнет от того, что несколько крупных банков проворовались, но такого рода событие может запустить эффект домино общей неустойчивости.

Ошибки, связанные с переоценкой или недооценкой значения морального состояния общества и его элит. Одна из версий крушения Древнеримской империи — деградация её элит, состоящая в том, что люди, из которых рекрутировались правители всех уровней, действовали исключительно в своих личных краткосрочных интересах, иначе говоря, глупо и эгоистично (что может быть связано с тем, что они употребляли воду из водопровода со свинцовыми трубами, отрицательно влиявшую на мозг). При этом предполагается, что эффективное действие в своих долгосрочных интересах совпадает с интересами общества в целом, что, вообще говоря, не бесспорно. Другой метафорой является сравнение «морального духа», например, войска — со способностью молекул некого вещества превращаться в единый кристалл (подробно на эту тему рассуждал Лев Толстой в «Войне и мире»).

С другой стороны, на падение нравов жаловались ещё сами древние римляне, и до сих пор этот процесс не помешал развитию производственных сил общества. Корень ошибки здесь может быть в конфликте поколений, а именно в том, что опытные и старые оценивают

молодых и задиристых, не внося возрастную поправку и забывая, что сами были такими же.

Ошибка, связанная с тем, что вместе того, чтобы исследовать истинность или ложность некого сообщения о риске, человек стремится доказать эту идею как можно большему числу людей [Yudkowsky, 2007]. Одни идеи проще доказывать, чем другие. Это приводит к сдвигу в оценке вероятностей.

Склонность людей предлагать «простые» и «очевидные» решения в сложных ситуациях — не подумав. [Yudkowsky, 2007]

А затем упорствовать, защищая их и подбирая под них аргументацию. Закон Мёрфи «Любая сложная проблема имеет простое, очевидное и неправильное решение».

Общественная дискуссия о рисках разных исследований может привести к тому, что учёные будут скрывать возможные риски, чтобы их проекты не закрыли. «*И если власти вводят закон, по которому даже мельчайший риск существованию человечества достаточен для того, чтобы закрыть проект; или если становится нормой де-факто политики, что ни одно возможное вычисление не может перевесить груз однажды высказанного предположения, то тогда ни один учёный не рискнёт больше высказывать предположения*». [Yudkowsky, 2007]

Преждевременные инвестиции. Если бы в середине XIX в. люди бы поняли, что в XX в. им угрожает атомное оружие, и на предотвращение этого риска были бы выделены миллионы, то нет сомнений, что эти деньги были бы потрачены не по назначению, и у будущих поколений выработалась бы аллергия на такие проекты. Возможный пример: по некоторым данным, СССР в 80-е гг. получил дезинформацию о том, что США во всю разрабатывают беспилотные летательные аппараты, и развернул свою огромную программу, в результате которой возникли такие аппараты как «Пчела» — автоматические самолёты-разведчики весом около тонны, огромной стоимости и малой надёжности. В результате российские военные разочаровались в дронах именно к тому моменту, когда в США была принята программа их реального создания.

Склонность людей путать свои ожидания того, как оно будет скорее всего, и того, как оно будет в наилучшем случае. «*Реальность, как оказалось, зачастую преподносит результаты, худшие, чем самый наи-*

худший случай» [Юдковский, 2006]. Юдковский описывает в своей статье эксперимент со студентами, где их просили оценить наиболее вероятное и наихудшее время сдачи дипломной работы. В результате среднее время сдачи дипломной работы оказалось хуже, чем наихудший случай.

Апатия прохожего. Глобальные риски не являются чей-то личной ответственностью, и принято рассуждать в том смысле, что раз никто ничего не делает в связи с этим, то почему именно я должен? Более того, это состояние возникает бессознательно, просто как рефлекс подражания группе. Типичный пример: когда человек лежит на тротуаре и мимо идёт толпа, никто не помогает ему. Но если один человек на тропинке в лесу увидит лежащего человека, он, скорее всего, ему поможет. См. подробнее [Yudkowsky, 2007]

Потребность в завершении. Концепция когнитивной психологии, обозначающая стремление человека как можно скорее найти ответ на беспокоящий вопрос (need for closure — [Kruglanski, 1989]). Это приводит к тому, что человек предпочитает быстрое и неверное решение более долгому поиску правильного ответа. И хотя мы не можем искать правильную стратегию работы с глобальными рисками бесконечно долго — мы ограничены во времени! — нам стоит хорошо подумать перед тем, как придти к каким-то выводам.

Представление о том, что изменять обстоятельства следует, уничтожая их причины. Однако спичка, от которой загорелся пожар, уже погасла. Стремление уничтожить любую систему, от государства до тараканов и микроорганизмов, приводит к тому, что эта система оптимизируется для борьбы, становится сильнее. А тот, кто с ней борется, вынужден приобретать качества своего врага, чтобы действовать с ним на одной территории.

Забвение основного принципа медицины — «Не навреди!»

Путаница между объективными и субъективными врагами. Например, если встать перед мчащимся поездом и сказать, что поезд хочет меня уничтожить.

Страх утраты идентичности. Система не хочет глубоко трансформироваться, так как после трансформации это будет уже другая система.

Понятная катастрофа может быть привлекательнее непонятного будущего.

Склонность переоценивать ущерб от малых и частых аварий и недооценивать большие и редкие. Подробнее см. Юдковского [Yudkowsky, 2007].

Ошибка, связанная с неверным переносом закономерностей верных для одной системы в другую.

1) **Игнорирование роста усложнения структуры как фактора, снижающего надёжность системы.** Если от растения можно отрезать большую часть, не повредив его способности к полноценному восстановлению, то чтобы убить животное, достаточно удалить очень маленький кусочек организма. То есть, чем сложнее система, тем больше в ней уязвимых точек. Нельзя не отметить, что по мере процессов глобализации, связность и структурность земной цивилизации растёт.
2) **Снижение надёжности системы пропорционально четвёртой степени плотности энергии**. Это эмпирическое обобщение (точное значение степенного показателя может отличаться в зависимости от разных факторов) проявляется в сравнении надёжности самолётов и ракет: при равных размерах, количестве и затратах надёжность ракет примерно в 10 млн раз меньше — за счёт того, что плотность энергии в двигателях в несколько раз больше, и ряда других факторов. Похожее эмпирическое обобщение верно и для статистики аварий автомобилей со смертельным исходом в зависимости от скорости. Нельзя не отметить, что энерговооружённость человечества постоянно растёт.

Двусмысленность (многозначность) любого высказывания как источник возможной ошибки. С точки зрения авторов регламента по Чернобыльскому реактору персонал нарушил их требования, а с точки зрения персонала, пользовавшегося этим регламентом, они сделали всё в соответствии с этим регламентом. Регламент требовал «заглушить реактор» — но разработчики считали, что это должно произойти немедленно, а операторы — что постепенно. Другой пример — автоматическая система спасения и пилот могут совершать набор действий, каждое из которых в отдельности спасло бы самолёт, но вместе они накладываются друг на друга и приводят к катастрофе (гибель А310 в 1994 г. в Сибири).

Отказ рассматривать некий сценарий по причине его «невероятности». Однако большинство катастроф случаются в результате именно невероятного стечения обстоятельств. Гибель «Титаника» связана с уникальной комбинацией 24 (!) обстоятельств.

Общелогические ошибки, могущие проявиться в рассуждениях о рисках

Путаница между вероятностью, как мерой изменчивости объекта, и степенью уверенности, как мерой информации об объекте. Первое относится к вероятностному процессу, например, радиоактивному распаду, а второе к неизвестному процессу — например, угадыванию карты. Однако глобальные риски относятся к процессам, где мы вынуждены высказывать вероятностные суждения о процессах, который одновременно и вероятностные, и неизвестные. Здесь мы начинаем говорить о степени уверенности в той или иной вероятности. В этом случае вероятность и степень уверенности перемножаются.

Подмена анализа возможностей анализом целей. Например, суждение: «террористы никогда не захотят применять бактериологическое оружие, потому что оно нанесёт удар и по тем, чьи интересы они защищают». Структура целей может быть очень сложна или просто содержать в себе ошибки.

Неверное употребление индуктивной логики следующего вида: раз нечто очень давно не происходило, то это не будет происходить ещё очень долго. (Подробнее см. статью «Природные катастрофы и антропный принцип» в этом Сборнике)

Мышление, обусловленное желаниями — то, что по-английски называется «Wishful thinking». В зависимости от того, что человек хочет доказать, он будет фильтровать аргументы, часто неосознанно.

Логическая ошибка, в связи с попытками доказать, что нужно делать, исходя только из описания фактов. Однако если в первой и второй посылке умозаключения содержатся только факты, то и в выводе могут быть только факты. Любое рассуждение о целях должно опираться на некие представления о ценностях, заданных аксиоматически. Однако это означает произвольность таких целей, и их понимание может различаться у разных исследователей глобальных рисков, что может вести к разным определениям катастрофы и представлениям о том, что будет из неё выходом. Кроме того, любая система аксиом позволяет формулировать недоказуемые высказывания (теорема о неполноте), и в отношении долженствований в этом легко убедиться: почти любая система базовых ценностей легко позволяет создавать внутри себя противоречия, что является основ-

ным содержанием многих литературных произведений, где герой должен сделать выбор между, допустим, любовью к семье и к родине (то, что ещё называется экзистенциальный выбор). Неизвестно, возможна ли вообще непротиворечивая система ценностей, как она будет выглядеть, и будет ли применима на практике. Однако работа над непротиворечивой системой ценностей важна, так как её нужно будет вложить в будущие машины, обладающие искусственным интеллектом.

Ошибки, связанные с подменой анализа рисков анализом коммерческих мотивов тех, кто о них говорит. Можно рассуждать следующим образом: если человек исследует риски бесплатно, то он не вполне нормален, если он хочет получать за это деньги, то он паразитирует на общественных страхах, если это его прямые должностные обязанности, то доверять ему нельзя, потому что он запудривает мозги населению. Отсюда видно, что прямой связи между деньгами и анализом рисков нет, хотя в некоторых случаях она возможна. Объяснение через упрощение называется «редукционизмом» и позволяет объяснить всё, что угодно.

Использование так называемого «авторитетного знания». Ссылки на мнения великих людей не должны служить достаточным основанием, чтобы признать нечто безопасным.

Неправильное применение идеи о том, что теория должна быть опровержимой, чтобы быть истинной. Если рассматривать научный метод, как способ получения наиболее достоверных знаний, то эта методология должна быть верной. Однако с точки зрения обеспечения безопасности необходим противоположный подход: теория о существовании рисков является верной до тех пор, пока она не опровергнута.

Восприятие новой информации через призму старой. Есть данные, что только 7 % информации человек берёт из внешнего мира, а остальные додумывает сам. Увы, тоже самое верно и для текстов, в том числе и по глобальным рискам. Читая рецензии разных людей на один и тот же текст, не трудно убедиться, что они восприняли его совершенно по-разному. Вряд ли это связано с тем, что одни люди были принципиально умнее других — скорее, с тем, что они применяли разные фильтры восприятия. Более того, если человек начал придерживаться некой точки зрения, то он подписывается на те издания и выбирает те статьи, которые её подтверждают. Таким образом, у него создаётся иллюзия, что статистика по данным, подтверждающим его точку зрения, растёт. Это ещё более укрепляет его в своей правоте.

Ошибка в выборе нейтральной позиции. Каждый человек со временем понимает, что он не вполне объективен, и его точка зрения имеет некоторую тенденциозность. Чтобы компенсировать это отклонение, он может выбрать некий нейтральный источник информации. Ошибка состоит в том, что люди, придерживающиеся противоположных взглядов, выберут разные нейтральные точки, каждая из которых будет ближе к позиции того, кто её выбрал.

Уверенность как источник ошибок. Чем больше человек сомневается в своей точке зрения, тем чаще он меняет её под влиянием новых фактов, и тем больше шансов, что он попадёт к более достоверному знанию.

Истина не рождается в спорах о глобальных рисках. Споры способствуют поляризации точек зрения спорящих, каждый из которых начинает подбирать аргументы только в одну сторону, и уходит от состояния открытости ко всем остальным.

Употребление полностью дефектной логики. Увы, возможна ситуация, когда человек в своих рассуждениях совершает ошибки в каждой строчке. В этом случае он не мог бы найти свои ошибки, даже если бы хотел. Это может быть или одна повторяющаяся систематическая ошибка, или такая плотность разных ошибок, которая делает невозможным безошибочное рассуждении. Даже я сейчас не знаю наверняка, не делаю ли я каких либо систематических логических ошибок в настоящий момент. Это может происходить чаще, чем мы думаем — анализ текстов показал, что обычно люди пользуются сокращёнными умозаключениями и приёмами эвристики — и не осознают этого.

Различие между преднаукой и псевдонаукой. В тот момент, когда гипотеза находится в процессе формулирования, она ещё не обросла всем научным аппаратом и является скорее продуктом мозгового штурма на некую тему, возможно, осуществляемого коллективно путём обмена мнениями в печатных изданиях. И в этот момент она является преднаукой — однако она нацелена на то, чтобы стать частью науки, т. е. пройти соответствующий отбор и быть принятой или отвергнутой. Псевдонаука может имитировать все атрибуты научности — звания, ссылки, математический аппарат, — тем не менее, её цель — не поиск достоверного знания, а видимость достоверности. Все высказывания о глобальных рисках являются гипотезами, которые мы почти никогда не сможем проверить. Однако мы не должны отбрасывать их на ранних фазах созревания. Иначе говоря,

фаза мозгового штурма и фаза критического отсева не должны смешиваться — хотя обе должны присутствовать.

Ошибка, связанная с неправильным определением статуса «универсалий». Проблема универсалий была основной в средневековой философии, и состояла она в вопросе, что на самом деле реально существует. Существуют ли, например, птицы вообще, или существуют только отдельные экземпляры птиц, а все виды, рода и семейства птиц — не более чем условная выдумка человеческого разума? Одним из возможных ответов является то, что объективно существует наша способность различать птиц и не-птиц. Более того, каждая птица тоже обладает этим качеством. В рассуждения о рисках неясность по поводу универсалий вкрадывается следующим образом: свойства одного объекта переносятся на некий класс, как если бы этот класс был объектом. Тогда возникают рассуждения вроде «Америка хочет …» или «русским свойственно …», тогда как за этими понятиями стоит не единый объект, а множество, точное определение которого зависит от самого наблюдателя. Любые дискуссии о политике отравлены такого рода сдвигом. Рассуждая об искусственном интеллекте легко совершить такую ошибку, так как не понятно, идёт ли речь об одном устройстве или о классе.

Утверждения о возможности чего-то и невозможности неравносильны. Утверждение о невозможности гораздо сильнее, ибо относится ко всему множеству потенциальных объектов, а для истинности утверждения о возможности достаточно одного объекта. Поэтому большинство утверждений о невозможности чего-либо являются ложными гораздо чаще. Предполагая какое-то событие или стечение обстоятельств невозможным, мы наносим ущерб нашей безопасности. В определённых обстоятельствах возможно всё. При этом любые дискуссии о будущих катастрофах — это всегда дискуссии о возможностях.

Очевидности как источник ошибок. Правильное умозаключение всегда опирается на две посылки, два истинных суждения. Однако анализ текстов показывает, что люди очень редко употребляют полную форму умозаключений, а вместо этого употребляют сокращённую, где явно называется только одна посылка, а другая подразумевается по умолчанию. Умалчиваются обычно очевидности — суждения, кажущиеся настолько истинными и несомненными, что нет нужды их озвучивать. Более того, часто они настолько очевидны, что не осознаются. Понятно, что такое положение дел является причиной многочисленных ошибок, потому что оче-

видность — не обязательно истинность, и то, что очевидно одному, не очевидно другому.

Недооценка возможности ошибок в собственных суждениях. Я ничего не могу знать на 100 %, потому что надёжность моего мозга не равна 100 %. Я могу проверить свою надёжность, решив серию логических задач средний сложности, и затем посчитав количество ошибок. Думаю, будет меньше 95.

Ошибка, связанная с представлением о том, что каждое событие имеет одну причину. В действительности:

1) есть совершенно случайные события;
2) каждое событие имеет много причин (стакан упал, потому что его поставили с краю, потому что он сделан из стекла, потому что сила тяготения велика, потому что пол твёрдый, потому что кошка непослушная, потому что это рано или поздно должно было случиться);
3) каждая причина имеет свою причину, в результате чего мы имеем расходящееся в прошлое древо причин. Человеческий ум неспособен целиком это древо причин охватить и вынужден упрощать. Но понятие «причина» необходимо в обществе, потому что связано с виной, наказанием и свободой воли. То есть здесь под «причинной» имеется в виду принятие свободным вменяемым человеком решения о совершении преступления. Нет нужды говорить о том, сколько здесь неочевидных моментов. (Основной вопрос: Кто виноват?)

И в конструировании техники: где важно найти причину аварии. То есть то, что можно устранить — так чтобы аварий такого рода больше не было. (Основной вопрос: Что делать?)

Понятие «причины» менее всего применимо к анализу сложных уникальных явлений, таких как человеческое поведение и история. Пример тому — масса запутанных дискуссий о причинах тех или иных исторических событий. Именно поэтому рассуждения вроде «причиной глобальной катастрофы будет Х» — мягко говоря, несовершенны.

Ошибка, связанная с сознательным и бессознательным нежеланием людей признать свою вину и масштаб катастрофы. И вытекающие из этого неправильное информирование начальства о ситуации. Сознательное — когда военные скрывают некую аварию, чтобы их не наказали, желая справиться своими силами. Когда люди не вызывают пожарных, сами туша пожар до тех пор, пока не могут с ним справиться. Бессозна-

тельная — когда люди верят в то описание, которое уменьшает масштаб аварии и их вину. В Чернобыле организатор эксперимента Дятлов верил, что взорвался не реактор, а зал систем управления — и продолжал подавать команды на реактор. Вероятно, такое нежелание может действовать и вперёд во времени, заставляя людей не принимать на себя ответственность за будущие глобальные катастрофы.

Любопытство может оказаться сильнее страха смерти. Были жертвы в толпе любопытных, наблюдавших штурм Белого дома в 1993 г.

Система и регламент. Для оптимизации функционирования системы приходится позволять нарушать регламент по мелочам. Дальше действует следующая схема (описана в интернет-форуме по Чернобыльской катастрофе): «Мне приходилось принимать участие в расследованиях (или изучать материалы) несчастных случаев и аварий в промышленности (неатомной). По их результатам я для себя сделал следующий вывод: практически никогда не бывает какой-то „единственной главной“ причины и соответственно „главного виновного“ я имею в виду не официальные выводы комиссий, а фактическую сторону дела). Как правило, происходит то, что я для себя условно называю: десять маленьких разгильдяйств. Все эти маленькие разгильдяйства совершаются у всех на виду в течение многих лет подряд, а т. к. по отдельности каждое из них не способно привести к тяжелым последствиям, то в силу этого внимание на них не обращается. Но когда все они происходят в одно время, в одном месте и с одними людьми — это приводит к трагическому результату. Ну а когда происшествие имеет общественный резонанс — тогда обычно и назначают главного стрелочника по принципу: „кто не спрятался, я не виноват“».

Эффект «стрелочника». Вместо поиска подлинных причин аварии ищут стрелочника, в результате чего подлинные причины не устраняются, и она становится возможной ещё раз.

Необходимость выбора «актом веры». Если руководитель получает несколько противоречащих друг другу заключений о безопасности, то он делает выбор между ними, просто веря в одно из них — по причинам, не связанным с самой логикой.

Эффект первой и последней прочитанной книги. Порядок поступления информации влияет на её оценку, причём выделены первый и последний источник.

Преувеличение роли компьютерного моделирования. Наиболее две проработанные модели — метеорология и атомные взрывы. Обе составлены на огромном фактическом материале, с учётом сотен испытаний, которые вносили поправки к прогнозам, и обе регулярно давали ошибки. Даже самая точная модель остаётся моделью, т. е. не способна учесть все возможные случайности и детали происходящих явлений.

Доказательство по аналогии. Дело не только в том, что не может быть аналогий уникальному событию, которое ещё никогда не случалось — необратимой глобальной катастрофе, но и в том, что мы не знаем, как проводить такие аналогии. В любом случае, аналогии могут использоваться только в качестве иллюстраций.

Заключение

Масштаб влияния ошибок на рассуждения о глобальных рисках можно оценить, сравнив мнения разных экспертов, учёных и политиков по вопросу о возможности окончательной глобальной катастрофы и её возможных причин. Нетрудно убедиться, что разброс мнений огромен. Одни считают суммарные риски ничтожными, другие уверенны в неизбежности гибели человечества. В качестве возможных причин называется множество разных технологий и сценариев, причём у каждого будет свой набор возможных сценариев и набор невозможных сценариев.

Очевидно, что корни такого разброса мнений — в разнообразии движения мысли, которое, в отсутствии какой-либо зримой точки отсчёта, оказывается подвержено различным предубеждениям и ошибкам. Поскольку мы не можем найти точку отсчёта относительно глобальных рисков в эксперименте, представляется желательным, чтобы такой точкой отсчёта стала открытая научная дискуссия о методологии исследования глобальных рисков, на основании которой могла бы быть сформирована единая и общепризнанная картина глобальных рисков.

Литература

1. *Воробьёв Ю. Л, Малинецкий Г. Г., Махутов Н. А.* Управление риском и устойчивое развитие. Человеческое измерение // Общественные Науки и Современность, 2000, № 6.
2. *Корнилова Т. В*. Риск и мышление // Психологический журнал, 1994. № 4.

3. *Корнилова Т. В.* Психология риска и принятия решений (учебное пособие). М.: Аспект Пресс, 2003.
4. *Корнилова Т. В.* Мотивация и интуиция в регуляции вербальных прогнозов при принятии решений // Психологический журнал, 2006. № 2 (Совместно с О. В. Степаносовой).
5. *Корнилова Т. В.* Многомерность фактора субъективного риска (в вербальных ситуациях принятия решений) // Психологический журнал, 1998. № 6.
6. *МакМаллин Р.* Практикум по когнитивной терапии: Пер. с англ. СПб.: Речь, 2001. 560 с. (Гл. Логические ошибки)
7. *Платонов А. В.* Восприятие риска в ситуациях, требующих принятия решения // Доклад на конференции «Lomonosov», МГУ, 1996.
8. *Тофлер, Элвин.* Шок будущего. Москва, АСТ, 2002
9. *Bostrom N.* Existential Risks: Analyzing Human Extinction Scenarios. Journal of Evolution and Technology, 9. 2001. (русский перевод: Ник Бостром. Угрозы существованию. Анализ сценариев человеческого вымирания и связанных опасностей. Пер. с англ. А. В. Турчина. http://www.proza.ru/texts/2007/04/04-210.html)
10. *Bostrom N. and Tegmark M.* How Unlikely is a Doomsday Catastrophe? Nature, Vol. 438, No. 7069, p. 754, 2005 (пер. с англ. А. В. Турчина: Макс Тегмарк и Ник Бостром. Насколько невероятна катастрофа судного дня? http://www.proza.ru/texts/2007/04/11-348.html)
11. *Dawes R. M.* Rational Choice in an Uncertain World. San Diego, CA: Harcourt, Brace, Jovanovich, 1988.
12. *Fetherstonhaugh D., Slovic P., Johnson S. and Friedrich J.* Insensitivity to the value of human life: A study of psychophysical numbing. Journal of Risk and Uncertainty, 14: 238–300. 1997.
13. *Kahneman D., Slovic P., and Tversky A., eds.* Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, 1982.
14. *Kahneman D. and Tversky* A. eds. Choices, Values, and Frames. Cambridge, U.K.: Cambridge University Press, 2000.
15. *Kruglanski A. W.* Lay Epistemics and Human Knowledge: Cognitive and Motivational Bases. 1989
16. *Posner Richard* A. Catastrophe: Risk and Response. Oxford University Press, 2004; vii + 322 pp
17. *Taleb N.* The Black Swan: Why Don't We Learn that We Don't Learn? New York: Random House, 2005
18. *Yudkowsky E.* Artificial Intelligence as a Positive and Negative Factor in Global Risk. Forthcoming in Global Catastrophic Risks, eds. Nick Bostrom and Milan Cirkovic, UK, Oxford University Press, to appear 2007 (русский перевод: Э. Юдковский. Искусственный интеллект как позитивный и негативный фактор глобального риска. Пер. с англ. А. В. Турчина http://www.proza.ru/texts/2007/03/22-285.html)
19. *Yudkowsky E.* Cognitive biases potentially affecting judgment of global risks. Forthcoming in Global Catastrophic Risks, eds. Nick Bostrom and Milan Cirkovic, UK, Oxford University Press, to appear 2007 (русский перевод: Э. Юдковский. Систематические ошибки в рассуждениях, потенциально влияющие на оценку глобальных рисков. http://www.proza.ru/texts/2007/03/08-62.html)
20. *Назаретян А. П.* Технология и психология: к концепции эволюционных кризисов // Общественные науки и современность, 1993. № 3. С. 82–93.