

# О РИСКАХ, СВЯЗАННЫХ С ОШИБКАМИ ЭКСПЕРТОВ И АНАЛИТИКОВ<sup>1</sup>

Н. А. Абрамова, С. В. Коврига

*Институт проблем управления им. В. А. Трапезникова РАН, г. Москва*

Представлен новый подход к анализу ситуаций, традиционно интерпретируемых как ошибки экспертов, аналитиков, лиц, принимающих решение. Подход опирается на анализ практической деятельности экспертов и теоретические модели, учитывающие как практику, так и психологические знания. Факторы риска, обусловленные неадекватной интерпретацией «ошибок», анализируются на известных примерах нарушения транзитивности предпочтений.

## ВВЕДЕНИЕ

В условиях расширяющегося внедрения в практику различных компьютерных технологий, обрабатывающих качественную информацию и оценочные данные, усиливается роль человеческих факторов, и становится актуальной проблема рисков, связанных с ними, особенно когда такие технологии применяются в ответственных приложениях с высокой ценой возможного ущерба [2]. Эта проблема относится к различным типам технологий поддержки организованной интеллектуальной деятельности (называются ли эти технологии информационными, экспертными, интеллектуальными, когнитивными, поддержки принятия решений или иначе)<sup>2</sup>. Ключевым признаком, объединяющим такие технологии, с точки зрения человеческих факторов и связанных с ними рисков, является формализация знаний экспертов, аналитиков, лиц, принимающих решение (ЛПР), и других носителей знаний в проблемной области. Для удобства анализа мы называем этих специалистов обобщенным понятием «эксперт-аналитик».

Наш анализ показывает, что одним из значимых аспектов проблемы рисков являются практические ситуации, в которых «носители» технологий оценивают знания экспертов-аналитиков в ходе их формализации и нередко квалифицируют их как ошибочные (см., например, работы [3, 4]). Альтернативное отношение к этим

ситуациям (не исключающее возможности ошибок) связано с понятием психологической корректности технологий и используемых в них методов и операций приобретения знаний у человека [2, 5, 6]. Интуитивный смысл этого понятия, введенного О. И. Ларичевым и его коллегами<sup>3</sup>, состоит в том, что человек может предоставлять недостаточно надежную информацию в условиях, когда разработчики тех или иных методов и технологий не учитывают психологических особенностей «человеческой подсистемы переработки информации»<sup>4</sup>.

Настоящая статья посвящена анализу рисков, связанных с такими ситуациями. Анализ основан на изучении практики применения экспертных методов и теоретических исследований, которые вкратце представлены далее. К ним относятся:

- модель действия стереотипов [2, 6], опирающаяся, в основном, на работу [7], а также наши исследования [8];
- обобщенная модель эксперта-аналитика;
- оригинальная концепция схем представления знаний, ориентированных на корректность [8, 9].

Статья завершается анализом рисков на примере известной ситуации нетранзитивности парных предпочтений.

<sup>3</sup> «Под психологически корректными мы понимаем такие операции получения информации от ЛПР, использование которых приводит к согласованным, непротиворечивым результатам» [5].

<sup>4</sup> Расширение понятия психологической корректности в работах [2, 6] рассчитано на учет факторов психологического влияния, которые действуют путем навязывания человеку не только отдельных операций, но и всей структуры знаний и деятельности, ведущей к получению конечного результата.

<sup>1</sup> Статья представляет собой расширенное изложение доклада, представленного на конференции [1], отредактированное и дополненное некоторыми ссылками на публикации.

<sup>2</sup> В докладе [1] такие технологии для краткости обобщенно называются интеллектуальными.



## 1. МОДЕЛЬ ДЕЙСТВИЯ СТЕРЕОТИПОВ И ФАКТОРОВ ВЛИЯНИЯ

Вкратце модель, определяющая действие стереотипов и других эвристик и факторов влияния при принятии решений [3, 4], такова.

- Всякая мыслительная операция, которая может быть интерпретирована как принятие решения<sup>5</sup>, осуществляется посредством (не обязательно осознаваемой) операции подбора эвристики: готового решения или способа, при помощи которого затем находится и принимается решение<sup>6</sup>.
  - Человек располагает эвристиками разной степени проверенности, освоенности, пригодности для конкретной ситуации. Среди них особое место занимают стереотипы, которые обладают повышенной активностью (связываемой с механизмом условного рефлекса), благодаря закреплению опытом успешного применения или влиянию, например, научных парадигм.
  - Применение немотивированных эвристик, неадекватных конкретной ситуации, является довольно типичным, и этому способствуют:
    - объективные и субъективные условия, такие как недостаток времени, информационная сложность ситуации, неопределенность и др.;
    - специфика действия стереотипов при подборе эвристик;
    - факторы веры — факторы влияния, повышающие субъективную уверенность в адекватности решений, помимо объективной адекватности, вплоть до отклонений от рационального поведения.
- Типичными факторами веры являются принципы Чалдини: принцип последовательности, срабатывающий в силу обязательности (выбранной позиции), принцип социального доказательства, принцип авторитета.

## 2. МОДЕЛЬ ЭКСПЕРТА-АНАЛИТИКА

Исходя из анализа практической деятельности экспертов-аналитиков, которая не всегда согласуется с нормативными моделями структуры их знаний и деятельности в различных технологиях, и, ориентируясь на анализ рисков, принята следующая модель эксперта-аналитика.

Специалист является экспертом-аналитиком в некоторой области профессиональной деятельности, если он:
 

- *обладает совокупностью эвристик*, рутинное применение которых в стереотипных для него ситуациях обеспечивает относительно успешное решение (общих или частных) задач, адекватных ситуации, по крайней

<sup>5</sup> Принятие решения понимается в психологическом, а не формально теоретическом смысле; при этом, например, любая единичная операция оценивания — это принятие решения.

<sup>6</sup> В составе эвристик могут быть и знания, приобретенные из опыта, не обязательно осознаваемые, и научные знания, которые проверены на практике, или, по крайней мере, истинны и значимы для практики, с точки зрения эксперта, и мыслительные средства манипулирования профессиональными знаниями.

мере, в условиях отсутствия факторов влияния, толкающих к отказу от применения собственных профессиональных и интеллектуальных знаний и умений;

— способен с более или менее высокой степенью успеха *отличать нестереотипные (и пограничные) ситуации* из своей области деятельности и — в случае ситуаций, для которых его стереотипные эвристики неадекватны, — *творчески решать задачи* на основе собственных эвристик или привлекаемых знаний, *ориентируясь на успешность конечных результатов* этой деятельности<sup>7</sup>.

Предполагается, что знание, которое формирует эксперт-аналитик, действующий в области стереотипных для него ситуаций, является наиболее достоверным из-за успешности используемых эвристик, подтвержденной практикой, тогда как в случае нестереотипных ситуаций достоверность снижается из-за риска неадекватных решений. (Для простоты мы не касаемся ситуаций, когда результаты работы эксперта-аналитика недостоверны из-за недостоверной исходной информации или иных объективных факторов).

Предполагается также, что за рамками названных областей более или менее результативной (по успешности, достоверности) деятельности, в которых человек выступает как эксперт-аналитик за счет репродуктивного или творческого переноса своих профессиональных знаний и умений, лежат области с меньшей достоверностью результатов, в которых возрастает роль немотивированных эвристик. Грубо говоря, они составляют область «обыденного мышления», когда эксперт уже перестает быть экспертом<sup>8</sup>.

Помимо характера решаемых задач и имеющихся для этого мыслительных средств, на достоверность результатов деятельности эксперта-аналитика в рамках определенной технологии влияет *фактор навязанности способностей мышления*, который обусловлен регламентацией структуры знаний и интеллектуальной деятельности эксперта-аналитика. В частности, речь идет о моделях знаний, в соответствии с которыми его знания (исходные или порождаемые в ходе анализа) должны быть формализованы для их последующей обработки формальными методами.

О влиянии фактора навязанности на достоверность результатов и на наличие в них ошибок прямо или косвенно свидетельствуют как практические наблюдения, так и теоретические исследования, в том числе, по психологической корректности интеллектуальных и экспертных технологий [2, 5, 6]; по действиям человека в условиях, когда алгоритм в него «введен извне» [11].

Как следует из экспериментальных исследований и объяснительных гипотез психолога С. И. Шапиро [11], фокусирование сознания на выполнении отдельных умственных действий угнетает в сознании «логический

<sup>7</sup> В психологии выделенные области деятельности соответствуют репродуктивному переносу стереотипных знаний и творческому переносу [8].

<sup>8</sup> Противопоставление экспертного и обыденного мышления заимствовано из работы Л. Г. Ионина [10], который различает эти виды мышления по используемым стереотипам (хотя и в несколько иной терминологии).

механизм», «генерирующий и контролирующий эти действия» и придающий им целостность. Затруднения возникают из-за недостаточного понимания роли отдельных действий в общей структуре; теряется «схватывание целого» (гештальта), которое играет решающую роль. Человек оказывается в ситуации «операторного» уровня мышления, характерного для неопытных программистов.

Изложенная модель эксперта-аналитика в сочетании с моделью действия стереотипов и факторов влияния позволяет предположить, что навязанность определенных форм (схем) мышления является значимым фактором в ряду условий, при которых те или иные методы решения задач, основанные на операциях получения знаний от человека, становятся психологически некорректными. Разумеется, степень риска снижения достоверности знаний, получаемых от эксперта-аналитика знаний при недостаточном использовании его собственных ресурсов компетентности, зависит от многих факторов, и этот вопрос требует изучения.

### 3. СХЕМЫ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ, ОРИЕНТИРОВАННЫЕ НА КОРРЕКТНОСТЬ

Чтобы связать модель эксперта-аналитика, в рамках которой рассматривается вопрос достоверности знаний, с рисками в ситуациях, которые расцениваются как ошибки носителей знаний о проблемной области, обратимся вначале к понятию корректности (отсутствия ошибок).

Наши прежние поиски ответа на вопросы о том, какие знания считать корректными и какими путями обеспечить их корректность, привели к гипотезе, согласно которой *человеческие способности оценивать те или иные знания как некорректные связаны с представлением знаний по определенным схемам*<sup>9</sup> [8, 9]. Возникла концепция *схем представления знаний, ориентированных на корректность*, т. е. схем, которые, благодаря присущей им общей структуре знаний или определенным семантическим ограничениям на наполнение этой структуры, обеспечивают определенные свойства корректности конкретных знаний<sup>10</sup>.

Поясним действие схем представления знаний, ориентированных на корректность, при формализации, т. е. представлении конкретных знаний по определенной схеме. (Предполагается, что схема является общей для соответствующего типа знаний.) При формализации или порождении новых знаний по заданной схеме воз-

<sup>9</sup> Известное понятие схемы представления знаний, или просто схемы, довольно подробно анализируется в работе [8].

<sup>10</sup> Эти вопросы, в первую очередь, исследовались в рамках проблемы надежности программного обеспечения, которая была порождена ошибками в компьютерных программах. Было известно, с одной стороны, что формализация способствует обнаружению ошибок, а с другой, — что она может быть источником новых ошибок, ошибок перевода по Г. Майерсу [12]. Вместе с тем уже тогда нам стало понятным, что проблема некорректности формализованных знаний может быть актуальной и для информационных систем и систем, основанных на знаниях [9].

можно конфликтная ситуация. Она состоит в том, что общая схема, по априорной оценке человека, применима к рассматриваемому конкретному знанию, но представить его по этой схеме, иначе говоря, вложить его в схему, не удастся: знание оказывается так или иначе не соответствующим схеме. Такой конфликт нередко выражается в форме субъективных оценок типа «не понимаю», «что-то тут не так», «неправильно» и т. п.<sup>11</sup>

Типичная реакция носителя знаний в случаях конфликта между схемой и конкретным знанием, которое в нее вкладывается, заключается в том, что *приоритет схемы над конкретным знанием признается как какая-то некорректность этого знания*, даже если в первичном представлении (т. е. в представлении по другим, не обязательно осознаваемым, схемам) некорректность не была видна. Именно в этом, в *признании приоритета общей схемы* (заимствованной из логики, опыта, определенной теории, действующих норм и т. д.) *над конкретными знаниями* при оценке их корректности и состоит *ключевое свойство схем, ориентированных на корректность*.

По-видимому, древнейшими общепризнанными схемами, ориентированными на корректность (для знаний в виде умозаключений), являются силлогизмы Аристотеля, которые выглядят присущими человеческому разуму. Характерными примерами некорректностей, выявляемых посредством подходящих схем, служат неполнота, неоднозначность, противоречивость и т. п. Оценка знания как некорректного, содержащего какую-то ошибку (или ошибки), означает признание необходимости его корректировки, например, дополнения недостающего, согласно схеме, фрагмента.

Формально субъективную реакцию носителя первичных знаний на оценку результата применения схемы, ориентированной на корректность, можно описать утверждением

$$\begin{aligned} & \text{применимо } (S, k) \rightarrow \\ & \rightarrow (\text{некорректно } (k, S) \rightarrow \text{некорректно } (k)), \end{aligned} \quad (1)$$

где  $S$  — схема (общая для однотипных, в формальном аспекте, знаний);  $k$  — конкретное знание; *применимо*  $(S, k)$  — оценка применимости схемы  $S$  к знанию  $k$ ; *некорректно*  $(k, S)$  — оценка несоответствия знания  $k$  схеме  $S$  (иначе, наличия конфликта между ними), возникшая при применении  $S$  к  $k$ , в какой бы словесной формулировке эта оценка ни выражалась; *некорректно*  $(k)$  — оценка некорректности знания  $k$  как такового, безотносительно к схеме. При этом все оценки считаются бинарными (да, нет).

Выражение (1) означает, что если человек признает применимость общей схемы к его конкретному знанию, он признает и некорректность этого знания, если она проявляется при применении к нему схемы.

Как следует из выражения (1) некорректность знания относительно схемы проявилась, единственная возможность не отнести эту некорректность на счет конкретного знания — это принять, что схема не применима к данному знанию. И действительно, наряду с типичной

<sup>11</sup> Для простоты мы не рассматриваем ситуацию, когда трудности обусловлены тем, что человек не владеет схемой.



реакцией носителя знаний на конфликт конкретного знания и схемы, на практике случается смена априорной, до применения схемы, оценки «схема применима к данному знанию» на апостериорную — «схема не применима». Тем самым признается приоритет конкретного знания — надо менять схему.

Стоит уточнить, что, как показывают наблюдения, в конкретных ситуациях, подразумевающих оценку формализованного знания, роль схем может не осознаваться, но схемы проступают более или менее явно при обосновании необходимости коррекции. Особенно это заметно в ситуациях, когда конфликт обнаружен не носителем знания, а другим человеком. То же относится и к ситуациям, когда оценка «неправильно» появляется без предварительно заданной схемы. (Человек подбирает нужную схему (норму), чтобы обосновать свою оценку «неправильно».)

В изложенной концепции схем представления знаний, ориентированных на корректность, неявно предполагается, что схема может выступать как ориентированная на корректность в рамках определенного человеческого сообщества, если она признается в этом качестве всеми членами сообщества как *общая норма*. Это является *еще одним ключевым свойством* схем представления знаний, ориентированных на корректность. Оно создает основу для достижения соглашения относительно корректности конкретных знаний или необходимости исправлений.

Позиция носителя знаний выделена потому, что именно его знания должны подвергаться коррекции в случае обнаружения конфликта. Коррекции подлежат не только фрагмент, который оценивается как некорректный, но и, возможно, все, что с ним связано в его внутренних представлениях.

В различных технологиях в качестве схем, по которым осуществляется представление знаний экспертов-аналитиков, выступают *общие модели знаний, лежащие в основе технологий*. В целях сокращения рисков, ведущих к недостаточной достоверности знаний, порождаемых по некоторой технологии, естественно стремиться к тому, чтобы используемые модели знаний были ориентированы на корректность. В противном случае может иметь место *эффект маскирования ошибок*. Комплекс вопросов, касающихся рисков, обусловленных таким маскированием, составляет самостоятельную проблему. Далее мы ограничиваемся рассмотрением рисков, связанных с противоположными ситуациями, когда конфликт между общей моделью знаний и формализуемым знанием проявляется, что может, но отнюдь не обязательно, свидетельствовать об ошибках.

#### 4. РИСКИ ПРИ РАЗРЕШЕНИИ КОНФЛИКТА СХЕМЫ И КОНКРЕТНОГО ЗНАНИЯ

Рассмотрим ситуацию разрешения конфликта между общей схемой представления знаний эксперта-аналитика, лежащей в основе некоторой технологии, и конкретным знанием конкретного эксперта-аналитика. При

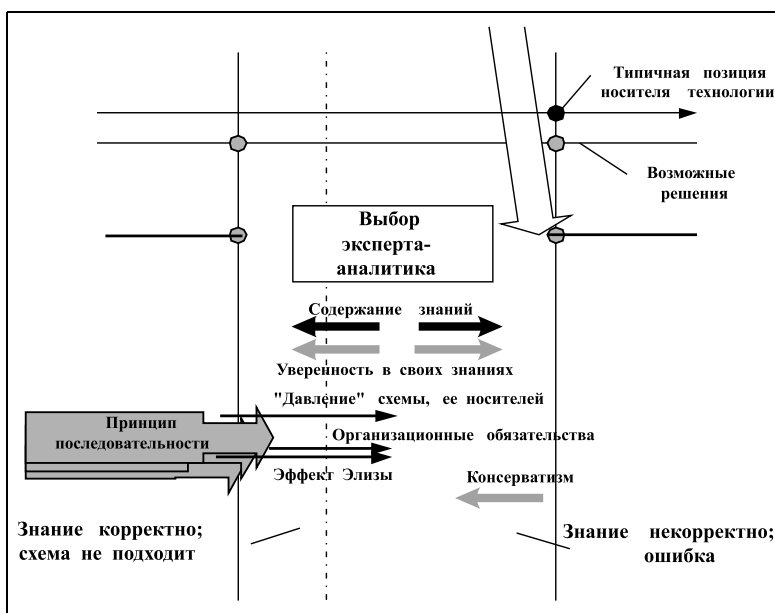


Рис. 1. Разрешение конфликта схемы и конкретного знания

этом не предполагается, что схема обязательно ориентирована на корректность. Предполагаем, что в процессе формирования формализованных знаний участвуют и эксперт-аналитик, и носитель технологии.

Имеются два возможных решения, снимающих конфликт. Одно из них соответствует оценке «знание некорректно; ошибка», другое — оценке «знание корректно; схема не подходит»<sup>12</sup> (рис. 1). Первое решение, в пользу схемы, означает необходимость коррекции знаний эксперта-аналитика, второе — отказ от схемы. Во втором случае при «жестких» технологиях возможен отказ от технологии или, напротив, от эксперта или его мнения.

Для носителя технологии типичная позиция состоит в решении «ошибка». Это легко объяснить моделью действия стереотипов: принятие приоритета схемы, заложенной в технологию, соответствует влиянию принципа последовательности. По той же модели, выбор эксперта-аналитика обусловлен целым рядом факторов влияния.

Вначале будем полагать, что решаемые задачи и формализуемые знания относятся к области наиболее достоверного знания эксперта-аналитика.

В сторону принятия адекватного решения, каким бы оно ни было, его толкают *объективное содержание знаний и субъективная уверенность в своих знаниях*. В сторону решения «ошибка», независимо от его адекватности, толкает, прежде всего, принцип последовательности, а также, в большей или меньшей степени — другие при-

<sup>12</sup> Для простоты мы исключаем из рассмотрения решения, требующие рефлексивного осознания того, что формализованное знание и первичное (внутреннее) знание эксперта-аналитика различаются по смыслу. Такое осознание требуется, в частности, для того, чтобы идентифицировать «описки» — расхождения между первичным знанием и его описанием.

нции и правила Чалдини: правило «взаимного обмена» (с носителями технологии), принципы благорасположения, авторитета, социального доказательства. Они воплощаются в ряде факторов, таких как «давление» схемы, которая уже принята и может быть более или менее признанной, организационные обязательства. Игрет роль и так называемый «эффект Элизы», известный по исследованиям психологов в области искусственного интеллекта, который состоит в преувеличенном доверии к решениям, полученным формальными методами. В своем большинстве факторы этой группы являются относительно объективными, т. е. не зависят от особенностей конкретной личности (хотя реакция на них зависит от личности). В противоположную сторону, своей правоты, независимо от адекватности решения, может толкать субъективный фактор консерватизма — неприятия навязанных схем.

Как показывает анализ выделенной совокупности факторов влияния, при разрешении конфликта *преобладающую роль играют систематические факторы влияния в сторону интерпретации конфликта как ошибки эксперта-аналитика*, прежде всего — принцип последовательности. При этом рассматриваемые факторы влияют лишь на субъективную уверенность в решении, *независимо от его объективной адекватности*.

Тем самым, *имеется риск ошибочно признанной ошибки* и, как следствие, ошибочно скорректированного знания эксперта-аналитика в ходе применения технологии: замены более или менее достоверного знания, не вкладываемого в схему, на результат его подгонки под схему. Объективная достоверность формализованных знаний при наличии такого риска снижается.

Основным противовесом от принятия экспертом-аналитиком ошибочного решения о признании ошибки служит его субъективная уверенность в своих знаниях, относящихся к области высокой достоверности его решений.

Специфика схем, ориентированных на корректность, при разрешении конфликта состоит, прежде всего, в существенном усилении влияния схемы, поскольку она признана как общая норма. Решая вопрос о рисках, связанных с признанием ошибки при использовании схем, ориентированных на корректность, естественно предположить, что такого рода схема становится нормой благодаря опыту успешного использования. Иными словами, схема, ориентированная на корректность, может рассматриваться как общепризнанная эвристика, ставшая нормой, подтвержденной опытом успешного использования. Как и всякая эвристика, она имеет свои границы успешного применения. Тогда можно ожидать, что риск мал или даже пренебрежим при использовании таких схем в своих границах. Можно ожидать увеличения риска в условиях, когда

— принятие схемы в качестве нормы не обусловлено опытом успешного использования;

— границы применимости нечетки.

Обратимся теперь к ситуациям, которые относятся к областям меньшей достоверности формализованных знаний. Обобщенно рассматривая изменения среди факторов, влияющих на принятие решения о наличии ошибки по отношению к ситуациям с высокой досто-

верностью решений, в среднем можно ожидать снижения субъективной уверенности эксперта-аналитика в достоверности формализованных знаний. (Например, это может происходить из-за осознания недостаточности своих профессиональных ресурсов для формирования знаний приемлемой достоверности.) Однако при этом нет оснований считать, что риск ошибочно признанной ошибки существенно изменится.

## 5. АНАЛИЗ ПРИМЕРА НЕТРАНЗИТИВНОСТИ ПРЕДПОЧТЕНИЙ

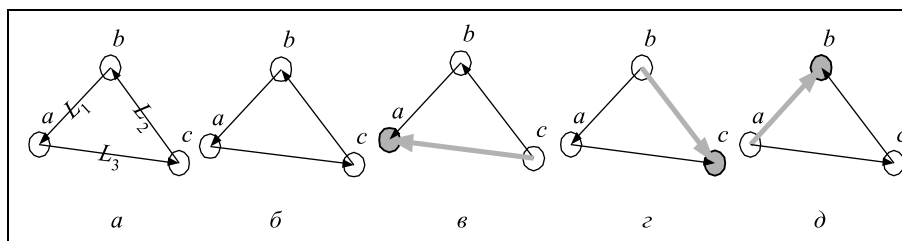
Рассмотрим ситуацию, которую обычно связывают с нарушением транзитивности строгих предпочтений, определяемых экспертом на некотором множестве объектов. Нетранзитивность отношения строгого предпочтения, «>», при ранжировании трех объектов  $a$ ,  $b$  и  $c$  проявляется в том, что имеют место оценки  $a > b$ ,  $b > c$  и при этом  $c > a$ , а не  $c < a$ , как это было бы в случае транзитивности. Нетранзитивность предпочтений довольно типична в практике принятия решений [4] и нередко интерпретируется сторонниками определенных технологий и их теоретиками как ошибка экспертов или, по крайней мере, нарушение тех или иных норм рациональности. Например, Ю. В. Сидельников говорит о «нарушении основного правила логического вывода — транзитивности» и непоследовательности экспертов [3], Б. Г. Литвак — о противоречивости и непоследовательности [4].

Если следовать логике теоретиков, говорящих о противоречивости и непоследовательности, и принять во внимание понятие психологической корректности операций получения информации от человека [5], операции парных сравнений, результат которых должен выражаться в виде строгих предпочтений, приходится рассматривать как психологически некорректные. Ведь признаками этого являются несогласованность и противоречивость результатов<sup>13</sup>.

Точка зрения, согласно которой применение принципа транзитивности является свидетельством рациональности мышления, а «нетранзитивное поведение» говорит об отклонениях от рациональности, распространена и среди психологов. Достаточно назвать широко известные работы А. Тверского и Д. Канемана [13].

Вместе с тем на сегодня накоплено немало свидетельств в пользу того, что применимость (адекватность) принципа транзитивности предпочтений ограничена. Исследования в этом направлении, свидетельствующие о динамике в понимании проблемы транзитивности, сегодня ведутся среди психологов. Назовем лишь недавнюю публикацию А. Н. Поддякова [14], в которой представлено сложившееся представление о принципе транзитивности предпочтений в психологии и собраны и обобщены результаты целого ряда исследований, опровергающих универсальность принципа. Опровержение, в основном, строится путем демонстрации объек-

<sup>13</sup> Нельзя не отметить, что сам О. И. Ларичев относил операции попарного сравнения к числу надежных, психологически корректных операций.



**Рис. 2. Варианты сравнительных оценок:**

$a$  и  $b$  — исходные экспертные оценки согласно отношениям (2) и (3),  $в$ ,  $г$  и  $д$  — альтернативные исправления «ошибки» в варианте 2 поведения эксперта

тивных отношений превосходства (как более или менее известных по публикациям, так и найденных автором), которые характеризуются нетранзитивностью.

Все больше свидетельств в пользу того, что нетранзитивные предпочтения могут иметь рациональные обоснования и объяснения, сегодня находят специалисты в области формальных методов принятия решений (см., например, работу [15]). Эти объяснения находятся в системе предпочтений экспертов, или, иначе говоря, в совокупности эвристик, применяемых при решении задачи. Характерная и, по-видимому, наиболее известная эвристика такого рода — это многокритериальная система предпочтений.

Хотя на сегодня ощутим недостаток единого методологического подхода к анализу многоаспектной проблемы транзитивности, который позволил бы согласовать различные точки зрения или, по крайней мере, объяснить их наличие и взаимосвязь, уже сейчас представляется целесообразным признать, что транзитивность предпочтений — это лишь проверенная практикой стереотипная эвристика. Она обеспечивает успешность на широком классе практических задач; однако, как и все эвристики, имеет ограниченную область применимости, о чем свидетельствует наличие многих контрпримеров.

Здесь мы остановимся на одном из аспектов нетранзитивности предпочтений: разрешении конфликтов при ее обнаружении в ходе применения экспертных методов и технологий и связанных с этим рисков. Анализ известного по публикациям примера, в котором применяются многокритериальные сравнения, проводится на основе представленных выше теоретических построений.

**Пример.** Пусть три объекта  $a$ ,  $b$  и  $c$  сравниваются экспертом на основе признаков (критериев)  $L_1$ ,  $L_2$  и  $L_3$ , и в каждой паре доминирующим при оценке оказывается один из признаков, при неразличимости (примерном равенстве)<sup>14</sup> по остальным. Результат сравнения можно представить отношением

$$(a \stackrel{L_1}{>} b \& a \stackrel{L_2}{\cong} b \& a \stackrel{L_3}{\cong} b) \& (b \stackrel{L_1}{\cong} c \& b \stackrel{L_2}{>} c \& b \stackrel{L_3}{\cong} c) \& (a \stackrel{L_1}{\cong} c \& a \stackrel{L_2}{\cong} c \& a \stackrel{L_3}{<} c),$$

которое, в свою очередь, исключением несущественных признаков сводится к отношению

$$a \stackrel{L_1}{>} b \& b \stackrel{L_2}{>} c \& a \stackrel{L_3}{<} c. \quad (2)$$

<sup>14</sup> Отношение примерного равенства не является транзитивным.

Тогда, при необходимости вложить этот результат экспертного анализа в схему строгих парных предпочтений, его естественно редуцировать к виду

$$a > b \& b > c \& a < c, \quad (3)$$

где обнаруживается нетранзитивность.

Здесь хорошо видно, что нетранзитивность получается благодаря применению к знаниям эксперта более грубой модели знаний, которая навязана технологией; эта схема требует обобщения результатов попарных сравнений в виде оценок предпочтения. Можно привести и другие примеры эвристик сравнения, приводящих к сходным последствиям.

Рассмотрим три варианта возможной реакции эксперта на оценку отношения (3) как ошибочного, в соответствии с требованием транзитивности, выполнение которого позволяло бы определить наилучший элемент. Отношения (2) и (3) графически представлены на рис. 2,  $a$  и  $b$ .

**Вариант 1.** Пусть эксперт нацелен на задачу определения наиболее предпочтительного элемента. Пусть при этом он уверен в адекватности своих эвристик, использованных при сравнении (в выборе критериев, в оценке по отдельным критериям), уверен в достоверности полученных результатов, и пусть он слабо подвержен влияниям в сторону признания ошибки. Он может даже априорно признавать транзитивность предпочтений (т. е., с формальной точки зрения, признавать, что отношение предпочтений является отношением строгого порядка). Проведя свой анализ, он может прийти к выводу, что требование транзитивности неприменимо к конкретной решаемой задаче, в силу установленного им отношения (2), которое означает отсутствие наиболее предпочтительного элемента и в котором он уверен.

Непризнание экспертом требования транзитивности для конкретной задачи рационально обосновывается не только его недостаточной мотивированностью, но и в силу факта отсутствия наилучшего элемента в конкретном примере.

**Вариант 2.** Пусть, решая ту же задачу, эксперт больше подвержен влиянию навязанной схемы или иным факторам влияния. Тогда при помощи отношения транзитивности, он легко может подогнать результат под требование корректности, выбирая за основу любую пару предпочтений. На рис. 2,  $в$ ,  $г$ ,  $д$  показаны три версии решения с использованием транзитивности для определения последней из трех оценок. (Серым цветом показано предпочтение, выводимое из более достоверных оценок, полученных по экспертной эвристике.) Недостоверность конечного результата: определения наиболее предпочтительного элемента, разного во всех трех вер-

сиях, — очевидна. Это — пример реализации риска исправления ошибочно признанной ошибки, результатом которой оказывается менее достоверное знание, чем исходное (рис. 2, а, в).

Понятно, что тем же путем эксперт мог действовать изначально, проводя детальное сравнение для каких-то двух пар элементов и к третьей применяя транзитивность. В сравнении с результатом более полного анализа всех трех пар это лишнее свидетельство о недостаточной мотивированности использования свойства транзитивности для решения задачи определения наиболее предпочтительного элемента при многокритериальном сравнении.

Если рассматривать отношение строгих предпочтений со свойством транзитивности в качестве схемы представления знаний эксперта при решении задачи определения наиболее предпочтительного элемента с позиций психологической корректности, даже рассмотренный частный пример позволяет считать, что эта схема обладает признаками психологической некорректности. Ведь она толкает эксперта к предоставлению знаний, не соответствующих его возможностям как эксперта (ресурсам компетентности). Правда, в данном случае некорректными оказываются не отдельные операции попарных сравнений (как это следовало бы из определения в работе [3]), а недостаточно мотивированное требование транзитивности системы предпочтений в целом (согласно определению из работы [6]). Источником риска можно считать то, что эвристика транзитивности опирается на обширный человеческий опыт успешного использования, сделавший ее стереотипом, при том, что границы ее применимости нечетки.

*Вариант 3.* Пусть эксперт нацелен на задачу (локальных) парных сравнений («от вас требуется только сравнение пар»), и при этом, как и в варианте 1, он уверен в адекватности своих эвристик и слабо подвержен влияниям в сторону признания ошибки. Оценка результата (3) (см. рис. 2, б) полученного редукцией адекватного отношения (2) (см. рис. 2, а) как ошибочного и здесь может быть не признана экспертом из-за неприменимости требования транзитивности, но уже по другой причине: транзитивность не требуется для определения предпочтения в любой отдельно взятой паре элементов. Последний вариант может показаться не убедительным «противной стороне». Однако следует иметь в виду, что задачи определения наиболее предпочтительного элемента и определения локальных предпочтений психологически различны и могут опираться на разные эвристики. Здесь мы вновь сталкиваемся с навязыванием психологически некорректного решения.

*Основные выводы,* которые можно сделать из анализа примера с формированием нетранзитивных экспертных предпочтений и гипотетических реакций экспертов на оценку «ошибка», в соответствии с описанным теоретическим подходом к анализу, таковы.

- Возможно более или менее рационально обоснованное непризнание экспертом требования транзитивности для конкретной задачи, причем — непризнание апостериорное (т. е. после применения к конкретным знаниям), опирающееся на приоритет конкретных знаний над общим требованием. Это является дополнительным свиде-

тельством того, что отношение предпочтений со свойством транзитивности не может рассматриваться в качестве общепринятой нормы или, иначе, в качестве универсальной схемы, ориентированной на корректность.

- Различные контраргументы в пользу транзитивности в случае многокритериальных сравнений, а значит, и в общем случае могут рассматриваться лишь как факторы психологического влияния на экспертов. (Анализ наиболее популярных контраргументов, таких как принцип «денежного насоса» А. Тверского или «нарушение основного правила логического вывода — транзитивности», выходит за рамки данной статьи, заслуживая отдельной публикации.)

- Можно ожидать, что некоторые ошибки, особенно относящиеся к области меньшей достоверности результатов деятельности эксперта, будут обнаруживаться посредством требования транзитивности (признаваться экспертами как ошибки); однако при этом имеется риск ошибочного признания ошибки.

- Из интуитивных соображений, в условиях таких эвристик сравнительной оценки, как многокритериальное сравнение, требование транзитивности агрегированных предпочтений представляется психологически некорректным, если оценивать его по достоверности результатов. Однако само понятие психологической корректности как теоретическое понятие требует развития и уточнений, особенно, — в контексте примеров, связанных с обнаружением «ошибок» и ошибок.

#### ЗАКЛЮЧЕНИЕ. ОБЩИЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ

- Концепция схем представления знаний, ориентированных на корректность, и примененный подход к анализу рисков при разрешении конфликта между общей схемой и конкретным профессиональным знанием, позволяют установить, что проявлениями риска могут быть ошибочно признанные ошибки и замена относительно достоверного знания на менее достоверный результат его подгонки под схему.

В свете этого можно считать, что недостаточная психологическая корректность определенных методов и технологий, поддерживающих деятельность по принятию решений, может выражаться не только в виде типичных ошибок, но и в маскировании недостоверности.

- Представленные качественные модели поведения и анализ рисков, проведенный на их основе, позволяют предположить следующее. Явление психологической некорректности при получении формализованных знаний от человека-носителя конкретных практических знаний (ЛПР, эксперта, аналитика), впервые обозначенное и исследованное О. И. Ларичевым и его коллегами, имеет более глубокие и многообразные корни, чем они, по-видимому, предполагали, учитывая лишь «ограничения человеческой системы переработки информации».

Представляется целесообразным проводить исследование этого явления и его роли в создании технологий компьютерной поддержки интеллектуальной деятельности человека с привлечением психологических знаний и психологов.



- Схема анализа факторов, влияющих на достоверность решения о признании экспертом-аналитиком своей ошибки, примененная в данной работе для обобщенного анализа, позволяет проводить анализ рисков в конкретных технологиях и даже отдельных ситуациях вплоть до учета особенностей людей, влияющих на принятие решения. Для такого анализа целесообразно применять рефлексивный подход, когда участники взаимодействия, носители знаний о проблемной области и конкретных решаемых проблемах и носители технологии взаимно оценивают друг друга.
- Ввиду практической значимости проблемы транзитивности предпочтений для решения многих практических задач управления представляется необходимым провести комплексный анализ проблемы с учетом всех сложившихся на сегодня теоретических точек зрения и известных фактов, включая и факт уверенности многих специалистов в универсальности принципа транзитивности, вопреки имеющимся знаниям.

## ЛИТЕРАТУРА

1. *Абрамова Н. А., Коверига С. В.* О рисках, связанных с ошибками экспертов и аналитиков // Тр. 4-й Междунар. конф. «Когнитивный анализ и управление развитием ситуаций». — М., 2004. — Т. 2. — С. 12—23.
2. *Абрамова Н. А.* О поиске подходов к созданию психологически корректных интеллектуальных технологий // Тр. 3-ей Междунар. конф. «Когнитивный анализ и управление развитием ситуаций» CASC'2003. — М., 2003. — Т. 2. — С.153—158.
3. *Сидельников Ю. В.* Теория и организация экспертного прогнозирования. — М.: ИМЭМО АН СССР, 1990.
4. *Литвак Б. Г.* Экспертные оценки и принятие решений. — М.: «Патент», 1996.
5. *Асанов А. А., Ларичев О. И.* Влияние надежности человеческой информации на результаты применения методов принятия решений // Автоматика и телемеханика. — 1999. — № 5. — С. 20—31.
6. *Абрамова Н. А.* О некоторых мифах в оценке качества программного обеспечения // Надежность. — 2004. — № 1. — С. 38—63.
7. *Чалдини Р.* Психология влияния. — СПб.: ПИТЕР, 2001. — 270 с.
8. *Поиск подходов к решению проблем / И. В. Прангишвили, Н. А. Абрамова, В. Ф. Спиридонов и др.* — М.: СИНТЕГ, 1999. — 192 с.
9. *Абрамова Н. А., Сперанская И. В.* Развиваемая семантическая модель как механизм формализации знаний // Тез. докл. Всес. конф. по искусственному интеллекту. — Переславль-Залесский, 1988. — Т. 1. — С. 44—48.
10. *Ионин Л. Г.* Понимание и экспертиза // Вопросы философии. — 1991. — № 10.
11. *Шапиро С. И.* Мышление человека и переработка информации ЭВМ. — М.: Сов. радио, 1980.
12. *Майерс Г.* Надежность программного обеспечения. — М.: Мир, 1980.
13. *Kahneman D., Tversky A.* Judgment under uncertainty: heuristics and biases. — Cambridge: Cambridge University Press, 1983.
14. *Поддьяков А. Н.* Непереходность (нетранзитивность) отношений превосходства и принятие решений // Психология. Журнал высшей школы экономики. — 2006. — Т. 3. — № 3. — С. 88—111.
15. *Новиков Д. А.* Современные проблемы теории управления организационными системами // В кн.: Человеческий фактор в управлении / Под ред. Н. А. Абрамовой, К. С. Гинсберга, Д. А. Новикова. — М.: КомКнига, 2006. — С. 313—344.

☎ (495) 334-92-09;

e-mail: [abramova@ipu.ru](mailto:abramova@ipu.ru)



## Новая книга

**Новиков Д. А. Теория управления организационными системами. — М.: МПСИ, 2005. — 584 с.**

Книга посвящена описанию основ математической теории управления организационными системами. Ее цель — показать возможность и целесообразность применения математических моделей для повышения эффективности функционирования организаций (предприятий, учреждений, фирм и т. д.).

Описаны более сорока типовых механизмов — процедур принятия управленческих решений (реализующих функции планирования, организации, стимулирования и контроля): управления составом и структурой организационных систем, институционального, мотивационного и информационного управления. Их совокупность может рассматриваться как «конструктор», элементы которого позволяют создавать эффективную систему управления организацией.

Адресована студентам вузов, аспирантам и специалистам (теоретикам и практикам) в области управления организационными системами.

Более подробную информацию можно найти на сайте [www.mtas.ru](http://www.mtas.ru).