

# О формальных моделях естественных рассуждений и их эпистемологической адекватности

Муса Плиев

МГУ им. Ломоносова, философский факультет  
кафедра логики

15 мая 2021

Формальные модели естественных рассуждений — сравнительно молодое направление в логике, зародившееся в конце прошлого века практически одновременно в исследованиях искусственного интеллекта и в философии. Важной особенностью таких рассуждений является ситуация недостатка знания для получения дедуктивных заключений. Другой особенностью можно назвать возможность нарушения принципа монотонности, присущего отношению классического следования: в процессе их осуществления может открываться новое знание, заставляющее агента отказаться от некоторых из полученных ранее заключений.

Принцип монотонности:

$$\Gamma \vdash \phi \implies \Gamma \cup \{A\} \vdash \phi$$

# Введение

Наконец, подобные модели иногда можно представить в аргументативной форме, некоторые аргументы могут давать противоречивые результаты, что вызывает необходимость в формулировке принципов предпочтения одного аргумента другому.

Среди широкого разнообразия контекстов, предполагающих подобного рода рассуждения, можно выделить оперирование с такими выражениями, которые Р. Рейтер в статье «логика для рассуждения по умолчанию» назвал высказываниями с «нечёткими (fuzzy) кванторами» — например, «большинство  $A$  есть  $B$ », «почти ни один  $A$  не есть  $B$ ». «Классический» подход к описанию рассуждений в таких контекстах предполагает рассмотрение множеств максимальных непротиворечивых предположений относительно имеющегося знания, причём в зависимости от нашей манеры рассуждения (скептической или наивной) мы можем вывести некоторые заключения из пересечения или объединения этих множеств соответственно.

Мы практически не будем касаться подобных формализаций, поскольку они требуют обширного введения в теорию, но отметим этот подход как существующий.

Мы можем также представить подобного рода рассуждения в форме направленных графов, узлы которых обозначали бы аргументы, а рёбра — отношения между аргумента (стоит отметить, что отношение следования лишь один из элементов множества отношений подтверждения). В таких случаях нам потребуются определённые правила для выбора предпочтительных аргументов.

# Введение

Нашей задачей в данном докладе будет описать теорию опровержимых рассуждений Поллока и с этих позиций дать некоторую оценку существующим подходам к описанию подобных кейсов. Выбор этой эпистемологии не случаен — она представляет достаточно чётко структуру аргументации, учитывает эмпирические данные и выражает множество различных деталей. Безусловно, она не лишена недостатков, мы постараемся их отмечать в процессе.

Джон Поллок предлагает описывать процесс рассуждения как процесс оперирования с обоснованными, но опровержимыми суждениями, что представляется обоснованным в условиях правдоподобных рассуждениях в ситуации нехватки знания. Для этого он вводит понятие основания (я буду называть также его причиной) (reason) для принятия того или иного убеждения. Основания в свою очередь могут быть опровержимыми (defeasible) и неопровержимыми (conclusive). К последним относятся посылки в умозаключениях по правилам классической логики.

# Основания и доводы против них

Среди первого рода оснований выделяется класс так называемых *prima facie reasons* — на основании этих причин мы принимаем некоторое суждение без дополнительных оснований (кстати, что характерно, сама возможность такого рода причин описывается в терминах высказываний с «нечёткими кванторами»). Для каждой из таких причин может существовать контрдовод (*defater*). Такие доводы против первичных оснований могут быть двух типов — первые, называемые *rebutting defeaters*, являются основаниями для принятия отрицания суждения, вторые ставят под вопрос обоснованность вынесения суждения, то есть связь между первичными причинами и принятым на их основании утверждением.

На основе данных понятий Поллок выстраивает структуру аргументации. Аргумент представлен как некоторый вывод из множества посылок в исчислении по некоторым правилам. Правила представляются интуитивно приемлемыми, кроме одного из них. Первые три разрешают введение в любой строке аргумента утверждения, лежащего в основе эпистемического базиса либо являющегося элементом множества посылок, а также введения формул, для принятия которых могут быть причинами предыдущие строки. Последнее правило представляет собой аналог теоремы дедукции. Она может иметь место во многих случаях, но в силу того, что разные условные высказывания обладают разными свойствами, был предложен также ослабленный вариант, предполагающий формализацию той связи, на которую направлены доводы, критикующие возможность принятия суждения.

# Аргументация

Что касается аргументов, то тут так же стоит отметить несколько важных понятий.

Во-первых, понятие противоречивых (self-deafeating) аргументов — таких, в которых в разных строках содержится некая формула и один из контрдоводов к ней (то есть, либо критика самого суждения, либо возможности его принятия на основании имеющихся причин).

# Аргументация

Что касается аргументов, то тут так же стоит отметить несколько важных понятий.

Во-вторых, это понятие аргументов  $n$ -ного порядка. По умолчанию, все непротиворечивые аргументы считаются аргументами нулевого порядка. Аргументы одного порядка могут быть причинами для отказа от суждений, содержащихся в других аргументах того же порядка. Все такие аргументы, для которых среди аргументов этого порядка не содержится причин для отказа от них, могут считаться аргументами следующего порядка. Неопровержимым считается такой аргумент  $A$   $m$ -того порядка, для которого выполняется следующее условие: для любого  $n$ , если  $n > m$  то не существует такого аргумента  $n$ -ного порядка, в одной из строк которого содержалась бы причина для отказа от аргумента  $A$ .

## Применительно к другим моделям

Итак, теперь можно вкратце представить себе картину «идеальной» модели рассуждений — она должна поддерживать формализацию выражений с «нечёткими кванторами», она должна быть способна формализовать оба типа контрдоводов относительно принятия суждений, и в неё должно быть возможно встроить какие-то адекватные принципы специфичности.

## И ещё о формальных моделях

Поллок также пишет об одном важном свойстве человеческой памяти, которую надо учитывать, строя модель естественной аргументации: когда человек приходит к некоторому убеждению, он не всегда бывает способен вспомнить, как оно было получено, но использует его в качестве основания для следующих утверждений. Это означает, что при отказе от некоторых суждений мы не автоматически отказываемся от полученных заключений и процесс отказа от них требует более сложной процедуры, чем предполагается большей частью формализаций, построенных на основе представления об агенте как о носителе множества знаний, замкнутого относительно отношения логического следования (как в AGM Model или известных формализациях типа логики с умолчаниями или немонотонной логики Мак-Дермотта и Дойла).

# Заключение

Предложить работающий вариант такой формализации было бы весьма заманчиво, хотя и представляется труднодостижимым. В этом состоит одна из главных сложностей подобных моделей (хотя существуют и другие, учёт которых сделал бы попытки дать адекватную формализацию сущим адом).

Спасибо за внимание!