

УДК 159.955.6

Когнитивные функции дилеммы в свете проблем искусственного интеллекта

Восковская Л.В.¹, Куликов Д.К.²

¹ Южный федеральный университет

² Донской государственный технический университет

Аннотация: В статье предпринят анализ когнитивных функций дилеммы, с целью показать на примере этой логической формы содержательный характер мышления и пределы формализации, лежащие в основании программ ИИ. Рассматриваются содержательные особенности сложной дилеммы и анализируются понятия решения и разрешения дилеммы, а также задача выбора между заключенными в ней альтернативами. Утверждается, что разрешение дилеммы есть формально-логический алгоритм вывода, обеспеченный аксиоматическим статусом дилеммы, а решение дилеммы подразумевает раскрытие заключенного в ней противоречия и поиск новых отношений между ее условиями и альтернативами. Обоснование этого показано на примере задачи из жизненного опыта Г.С. Альтшуллера. Рассматриваются особенности формализации моральных дилемм и значение внелогических факторов процесса их решения. Показаны ограничения программирования морально квалифицированных систем ИИ, работающих в системе координат ригористической этики и естественного права.

Ключевые слова: дилемма, противоречие, искусственный интеллект, мышление.

Несмотря на очевидные успехи в области разработки искусственного интеллекта (ИИ), ни одну из созданных на настоящий момент «думающих» машин, способных к пониманию речи и текста, вычисляющих сложные шахматные алгоритмы и т.д., нельзя по-настоящему назвать разумной. Сегодня ИИ успешно решает многие прикладные задачи, однако вопрос о принципиальном отличии разума человека от «мышления» машины остается не снятым с повестки дня. ИИ в науке понимают по-разному; однако применительно к человеческому мышлению, которое выступило аналогом для создания машинного интеллекта, он определяется в терминах формальной логики и вычислительной операциональности. Попытки запрограммировать работу человеческого ума показали недостаточность выделения элементарных операций для решения сложных задач. Одним из направлений психологических исследований в связи с данной проблемой

является изучение роли субъектных качеств мыслящего, определяющих содержание мышления и недоступных машине.

К числу сложных задач, подвергаемых формализации, относится дилемма, которая в качестве логической формы определяется как особый вид умозаключений с условными и дизъюнктивными посылками. Дилеммы бывают конструктивные (*modus ponens* дилеммы) и деструктивные (*modus tollens* дилеммы), которые в свою очередь делятся на простые и сложные. Отличие сложной дилеммы от простой в том, что в ее заключении утверждается альтернатива, при этом с содержательной точки зрения имеет значение исключаящий характер разделительной посылки. Цель статьи – показать на примере дилемм содержательный характер мышления и пределы формализации, лежащие в основе программирования ИИ.

В психологии дилеммы стали популярным инструментом для изучения ряда психических феноменов; при этом детального анализа дилеммы с учетом ее формально-логических характеристик проведено не было. Чаще всего научное значение дилемма приобретает в исследовании морального сознания личности, где под дилеммой понимается положение, в котором оказывается человек, когда ему надлежит сделать выбор между двумя равно неприятными для него возможностями. Актуальной задачей становится вскрытие внутренних психологических механизмов принятия морально значимых решений при выборе одной из предлагаемых дилеммой альтернатив. Полемика относительно доминанты рациональных либо эмоциональных механизмов принятия решения в случае моральной дилеммы привела и к возрастанию интереса к сфере морали со стороны информатики в области права. За рубежом обсуждаются вопросы о создании «электронного судьи» и систем машинной этики (робоэтики) и разработки вычислительной теории морали, проводятся конференции по компьютерным приложениям деонтических систем [5], [6].

Данные тенденции ставят перед исследователями ряд вопросов, касающихся специфики дилеммы и ее функций как формы интеллектуальной деятельности, а также определения возможностей алгоритмизации решения сложных дилемм и ограничений машинного интеллекта. Дилемма как логическая форма, с одной стороны, обладает конкретной структурой, что дает возможности для ее формализации, но с другой стороны она – специфически интеллектуальный продукт. Дилеммы – это умозаключения, тесно укорененные в практику содержательных доказательств, в частности они могут быть сведены к известному в логике научных рассуждений доказательству разбором случаев. Личностная «отнесенность» дилеммы характеризуется рядом свойств. Если формулы простых дилемм содержат в себе алгоритм своего решения, то сложные дилеммы формализуют лишь необходимую альтернативу в содержании требуемого решения. Однако формальное разрешение не является предметным и практическим решением, что и составляет суть заключенной в дилемме проблемы. Следовательно, стоит различать *разрешение* дилеммы и ее *решение*. О *разрешении* дилемм говорится в смысле их аксиоматического статуса логических законов, в соответствии с чем может быть доказан алгоритм вывода из соответствующих посылок (двух условных и разделительной). В отличие от этого критерии и способы *решения* дилеммы лежат за ее пределами.

Рассмотрим содержательные условия решения сложной конструктивной дилеммы. Они определяются, прежде всего, конкретными значениями истинности входящих переменных. Для дизъюнкции содержательная выполнимость в двузначной логике возможна в трех условиях. С учетом этого не решаемость дилеммы требует введения дополнительного основания, а именно одновременной невыполнимости консеквентов условных посылок, и строгой дизъюнктивной посылки.

Соответственно, содержательно строгая сложная дилемма должна иметь форму:

$$\frac{a \rightarrow (b \wedge \bar{d}), c \rightarrow (d \wedge \bar{b}), a \vee c}{b \vee d}. \quad (1)$$

В ином случае полагание a и c одновременно истинными ведет к противоречию в посылках: $(a \wedge c) \rightarrow ((b \wedge d) \wedge (\bar{b} \wedge \bar{d}))$. Дилемма по формуле (1) предметно не решаема в смысле невозможности примирить противоположности и разрешить заключенное в них противоречие.

Таким образом, главное в дилемме – это отношение альтернативы и ее противоречие. Существующее в дилемме противоречие может требовать либо форсированного выбора альтернатив, либо направленного волевого усилия по поиску решения. Выбор между альтернативами – это *выход из дилеммы* при том, что условия дилеммы продолжают действовать и определять существо дела. Здесь дилемма выступает лишь моментом субъективного представления условий и структуры предметной ситуации. При выборе одной из альтернатив сама дилемма не решается, а элиминируется. О решении сложной дилеммы может идти речь тогда, когда находится альтернатива, не заданная условными посылками, но при этом отвечающая требованию их обоих.

Примером может служить дилемма из жизненного опыта Г.С. Альтшуллера. Во время заключения в Лефортово Альтшуллера ставили на так называемый «конвейер»: всю ночь ведут допросы, но днем спать не дают. Понимая, что такой режим он долго не выдержит, Альтшуллер оказался перед лицом касалось бы неразрешимой проблемы: если спать, то будешь разбужен, а если не спать, то это приведет к истощению, соответственно, как сделать так, чтобы спать и не спать одновременно? Решением задачи стала идея оторвать два кусочка бумаги, нарисовать на них зрачки, закрепить их на веках и сесть в пределах видимости волчка. Сокамерник в это время создавал

видимость беседы. В сидячем положении и с закрытыми глазами Альтшуллер высыпался, тем самым разрешая противоречие между двумя формально и физически несовместимыми альтернативами [1].

Разрешение противоречия является одной из концептуальных идей ТРИЗ, разработанной Альтшуллером. В рассматриваемом примере автор стоит перед проблемой: спать днем нельзя, т.к. за ним следит надзиратель, и за сон в неположенное время могут последовать санкции. При этом спать в разрешенное тюремным режимом время – ночью – он тоже не может, т.к. его ведут на допрос. Выразим заключенную в ситуации дилемму на формальном языке логики:

$$((a \rightarrow b) \wedge (\bar{a} \rightarrow d) \wedge (\bar{b} \vee \bar{d})) \rightarrow (\bar{a} \vee a), \quad (2)$$

где a – я сплю; b – меня будят, \bar{a} – я не сплю, d – меня ждет истощение.

Альтернатива, предлагаемая дилеммой (2), вступает в противоречие с посылками. Если \bar{a} истинно, то d также истинно, но по условию дилеммы, недопустимо d , следовательно $\bar{\bar{a}}$, т.е. a . Истинность \bar{a} влечет истинность a . Т.е. верно, что $a \wedge \bar{a}$. Соответственно, поскольку \bar{b} (нежелательно b), то \bar{a} . Следовательно, альтернатива становится противоречием: $a \wedge \bar{a}$. Это следствие того, что противоречие имплицитно содержится условиях задачи: $((a \rightarrow b) \rightarrow \bar{a}) \rightarrow d$, хотя формально-логического противоречия в этой формуле нет. Противоречие связано с функциональной ролью опосредования импликации переменной b .

Здесь следует предвосхитить возможное возражение: задача Альтшуллера представляет собой дилемму лишь по виду, но в действительности в ней дана не разделительная посылка, а конъюнкция «не быть разбуженным и не довести себя до истощения» ($\bar{b} \wedge \bar{d}$). Идеально, т.е. с точки зрения цели, это так. Но действительная связь моментов задачи процессуальная, поэтому сама ее формулировка содержит дилемму. Если

нежелательно b , тогда «не спи». Но b , опосредованно через \bar{a} влечет d во времени, и если желательно \bar{d} , то «спи», что влечет \bar{a} .

Физический алгоритм ситуации Альтшуллера исключает дилемму (маскирует задачу). С формальной точки зрения не существенен способ, каким человеку создается препятствие ко сну. Это может быть действие надзирателя, либо регулярно повторяющийся звук и т.п. Соответственно, здесь как бы нет конъюнкции условно-гипотетических оснований, и $(a \rightarrow \bar{a}) \rightarrow d \vdash (\bar{a} \vee \bar{a}) \rightarrow d \vdash \bar{a} \rightarrow d$. Однако, возможность превратить эту ситуацию в дилемму оказывается условием для поиска решения. Прежде всего нужно понять, что формально несущественный компонент b (действие надзирателя) оказывается *существенным* в данной системе отношений, поскольку надзиратель реагирует на глаза заключенного (закрыты они, либо открыты). Условия задачи удерживают это обстоятельство в форме отрицания (\bar{b}). При каких условиях эта формула имеет значение истины? 1) В том случае, если b ложно, т.е. надзиратель не будит. Но это опровергает гипотетическое основание условий, т.к. делает при истинности a значение импликации $a \rightarrow b$ ложным. 2) В том случае, если ложно a , т.е. Альтшуллер не спит ($\bar{a} \rightarrow \bar{b}$). Следовательно, решение может быть найдено только на принятии истинности противоречия, заданного альтернативой: $(a \wedge \bar{a}) \rightarrow (\bar{b} \wedge \bar{d})$. Кстати, истинность этой импликации не зависит от логического значения antecedента самого по себе при практической осуществимости (истинности) консеквента. В действительности Альтшуллер спал, но в системе отношений «на конвейере» он не спал, т.к. надзиратель его не будил. Альтшуллер воспроизвел себя как не спящий в социальном отношении и спящий в физиологическом за счет удержания определенного поведения надзирателя. Анализ ситуации Альтшуллера показывает, что если условия решения дилеммы заключают формально-логическое противоречие, то формализация решения становится принципиально невозможной.

Аксиоматический статус сложных дилемм включает в себе лишь постановку задачи, т.е. начальную фазу мышления. В них доказана лишь форма вывода, указывающая либо на необходимость выбора, либо на поиск условия обоюдной истинности альтернатив. Действительность мышления обнаруживается лишь с этого момента. Также обнаруживается задача определения условий *объективной неразрешимости* дилеммы, когда для ее решения не существует необходимых средств, а их целенаправленное создание оказывается недоступным. Такие дилеммы нередки в исторических коллизиях, как например в описанной Ф. Энгельсом дилемме Т. Мюнцера. «Самым худшим из всего, что может предстоять вождю крайней партии, является вынужденная необходимость обладать властью в то время, когда движение еще недостаточно созрело для господства представляемого им класса и для проведения мер, обеспечивающих это господство. То, что он *может* сделать, зависит не от его воли, а от того уровня, которого достигли противоречия между различными классами, и от степени развития материальных условий жизни, отношений производства и обмена, которые всегда определяют и степень развития классовых противоречий. То, что он *должен* сделать, чего требует от него его собственная партия, зависит опять-таки не от него самого, но также и не от степени развития классовой борьбы и порождающих ее условий... Таким образом, он неизбежно оказывается перед неразрешимой дилеммой: то, что он *может* сделать, противоречит всем его прежним выступлениям, его принципам и непосредственным интересам его партии; а то, что он *должен* сделать, невыполнимо» [3, с. 422-423].

Поиск средств решения задач во многом есть такое преобразование условий, которое не может быть осуществлено лишь в умственном плане. Решение, не имеющее условий технического воплощения на практике, является неполным. Часто это результат недостатка знаний о сущности

предмета. Выделение существенного элемента в условиях задачи нередко оценивается как продукт проницательности и *интуитивного* проникновения в проблему.

Особого обсуждения в свете исследуемого вопроса о когнитивных функциях дилеммы достойна *проблема формализации моральных дилемм*. Американский психолог Л. Колберг применял разработанный им метод дилемм для выявления уровней развития морального самосознания личности. Суть этого метода в том, что испытуемый ставится перед лицом необходимости осуществить выбор альтернативных действий, отождествляя себя с участником предлагаемой ситуации: например, либо украсть лекарство ради спасения жизни близкого человека, либо остаться в рамках закона ценой жизни этого человека. В подобных дилеммах в сознании испытуемого обнажается сложность отношений различных аспектов действительности, благодаря которым он может рефлексировать над своими ценностями и их осуществлением посредством выбора поведения.

Анализ моральных дилемм показывает, что при решении дилеммы существенным моментом является не только внутреннее отношение альтернатив, но и то значение, какое эти альтернативы приобретают по отношению к действительности. Дж. Грин, анализируя механизм решения моральных дилемм, показал, что чем сильнее заключенный в моральной дилемме конфликт между альтернативами, тем больше человек склонен выбрать одну из возможностей нерелексивно, импульсивно и эмоционально [4]. Однако нравственное действие всегда сопряжено с его субъективной оценкой, а следовательно, содержит в себе момент мышления. Это ставит вопрос, можно ли считать нравственной оценкой определение поступка через приписывание ему значения в соответствии с формализованной системой координат «плохо», «хорошо», «нейтрально»? Правильнее, на наш взгляд, что нравственная оценка предполагает процесс выстраивания этих координат

применительно к каждой конкретной ситуации, в которой разворачивается действие, и, тем самым, содержит в себе требование ее предварительного понимания и оценки. Убийство, видимо, следует однозначно приписать к морально негативной форме поступка, однако исчерпывается ли его нравственное значение константой «плохо»? Очевидно, что негативной всеобщностью обладает ситуация, в которой убийство оказывается необходимым моментом: плохо, что общество вынуждено казнить, плохо, что приходится воевать, плохо, что существуют преступники, плохо, что приходится защищать свою жизнь и жизнь близких от посягательств. Убийство как поступок наделяется нравственным качеством через его место в системе отношений. Абстрагирование от конкретного содержания элиминирует мышление из нравственного действия и превращает его в безликую реакцию. Такое понимание нравственного поступка в принципе предполагает неформализуемость морального поведения, и связь его развития с активным и свободным характером взаимоотношений людей. Другая точка зрения, возводящая нравственность к ритуализованному поведению и канону, определяет нравственность как свод правил и законов, ригористически противопоставляемых формам поведения, однозначно определяемым как аморальные и безнравственные. Именно такое понимание обсуждается в контексте возможности программирования морально квалифицированных систем ИИ, работающих в системе координат ригористической этики и естественного права (см., например, [2]).

Познавательные возможности дилеммы заключаются в том, что ее решение всегда подразумевает работу не только ума, но и личности мыслящего в целом. Дилемма имеет свое значение только в контексте человеческого сознания и человеческих отношений. Как показал случай Альтшуллера, принципиальное значение имеет понимание человеком *логики поведения* другого человека, сопряженной с его *ментальными состояниями*.

В машинном мышлении дилемма представляет собой конечный этап движения «вычисления», «выведения», «разрешения». В психолого-познавательном плане дилемма отражает способ структурирования проблемной ситуации логическими средствами; здесь дилемма не результат, а лишь начало мышления. Дилемма представляет собой уже вполне формализованную проблемную ситуацию. В ней четко определяются альтернативы, но будучи с психологической точки зрения проблемной ситуацией она может развиваться в процессе рассуждения в задачу. Развитие дилеммы подразумевает раскрытие заключенного в ней противоречия и поиск новых отношений между ее условиями, действительным положением дел и состоянием средств для действия. Именно такая система координат обозначает когнитивную функцию дилемм и задает вектор в оценках перспектив воссоздания мышления на основе программ ИИ.

Литература:

1. *Альтшуллер Г.С., Верткин И.М.* Как стать гением: Жизненная стратегия творческой личности. – Мн.: Беларусь, 1994.
2. *Лобовиков В.О.* Единство алетических, мажоритарных, эпистемических, аксиологических, деонтических и утилитарных модальностей // Научный ежегодник Института философии и права Уральского отделения Российской академии наук. 2009. Вып. 9. С. 123–140.
3. *Энгельс Ф.* Крестьянская война в Германии. Соч. 2-е изд. Т.7.
4. *Greene J.* Dual-process morality and the personal/impersonal distinction // J. of Exp. Soc. Psychol. 2009 45(3): 581-4.
5. Deontic Logic in Computer Science // 9th International Conference, DEON 2008, Luxembourg, Luxembourg, July 15-18, 2008, Proceedings
6. *Pereira L. M., Saptawijaya A.* Modelling Morality with Prospective Logic // Anderson M. (ed.) Machine Ethics. – Cambridge University Press, 2010.



References:

1. Al'tshuller G.S., Vertkin I.M. Kak stat' geniem: Zhiznennaja strategija tvorcheskoj lichnosti. – Mn.: Belarus', 1994.
2. Lobovikov V.O. Edinstvo aleticheskikh, mazhoritarnyh, jepistemicheskikh, aksiologicheskikh, deonticheskikh i utilitarnyh modal'nostej // Nauchnyj ezhegodnik Instituta filosofii i prava Ural'skogo otdelenija Rossijskoj akademii nauk. 2009. Vyp. 9. S. 123–140.
3. Jengel's F. Krest'janskaja vojna v Germanii. Soch. 2-e izd. T.7.
4. Greene J. Dual-process morality and the personal/impersonal distinction // J. of Exp. Soc. Psychol. 2009 45(3): 581-4.
5. Deontic Logic in Computer Science // 9th International Conference, DEON 2008, Luxembourg, Luxembourg, July 15-18, 2008, Proceedings
6. Pereira L. M., Saptawijaya A. Modelling Morality with Prospective Logic // Anderson M. (ed.) Machine Ethics. – Cambridge University Press, 2010.